

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 101.03 Seconds
(without alignments)
1052.139 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLMRSRLVLL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

all 57%

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	334	6.5	1751	20 AAY13493	Truncated cellulases
2	323	6.3	782	12 AAR15625	Cellulase AE-1. A
3	301.5	5.9	1749	22 AAB72314	Glutamic acid rich
4	276	5.4	1426	20 AAY13492	Truncated cellulases
5	274	5.3	406	22 AAG27250	Novel human diago
6	261.5	5.1	531	18 AAW15238	Scaffoldin protein
7	261.5	5.1	1853	19 AAW43108	C. thermocellum ce
8	244	4.8	1664	19 AAW43106	C. thermocellum OI
9	229.5	4.5	950	22 ABB71513	Drosophila melanog
10	227.5	4.4	455	22 AAU44346	Propionibacterium
11	225.5	4.4	551	18 AAW18790	Corrected Bacillus

12	219	4.3	1237	21 AAY81609	Streptococcus pneu
13	218	4.2	88	22 AAC02124	Human polypeptide
14	216	4.2	237	22 AAB74609	Leishmania major P
15	213	4.1	1357	22 ABB70775	Drosophila melanog
16	213	4.1	2344	22 AAU37120	Staphylococcus aur
17	212.5	4.1	521	22 AAB48788	Acidothermus cellu
18	212.5	4.1	598	22 AEG14000	Novel human diago
19	212	4.1	1532	21 AAB40945	Human ORFX ORF709
20	211.5	4.1	521	22 AAB48787	Acidothermus cellu
21	210.5	4.1	521	17 AAR89927	A. cellulolyticus
22	210.5	4.1	521	22 AAB48786	Acidothermus cellu
23	210.5	4.1	562	21 AAY69508	Acidothermus cellu
24	209	4.1	918	20 AAY08640	S. aureus ClfB pro
25	208.5	4.1	206	18 AAW14574	Streptococcus pneu
26	208.5	4.1	493	20 AAY28850	pectate lyase-link
27	208.5	4.1	493	20 AAY43218	pectate lyase CBD
28	207.5	4.0	1291	19 AAW59912	Amino acid sequenc
29	206.5	4.0	875	19 AAW34987	Bankia gouldi endo
30	206	4.0	1805	22 ABB65262	Drosophila melanog
31	205	4.0	311	22 AAU51992	Propionibacterium
32	205	4.0	1180	22 ABB65175	Drosophila melanog
33	204.5	4.0	499	14 AAR42122	NK-1 cellulase. B
34	204.5	4.0	957	21 AAY59288	Human MUC11 polype
35	204.5	4.0	957	22 AAM24513	C900P predicted am
36	202.5	3.9	476	21 AAY54123	A mannanase-linker
37	202	3.9	731	22 AAB74209	Protein encoded by
38	200.5	3.9	700	12 AAR13227	Novel endoglucanas
39	200	3.9	183	18 AAW14570	Streptococcus pneu
40	200	3.9	2586	22 ABB66878	Drosophila melanog
41	199.5	3.9	490	20 AAW95602	Saccharothrix aust
42	199.5	3.9	616	20 AAY13494	Truncated cellulase
43	199	3.9	317	21 AAG08664	Arabidopsis thalia
44	199	3.9	333	21 AAG46507	Arabidopsis thalia
45	199	3.9	448	21 AAB53201	Macaca mulatta tha

ALIGNMENTS

RESULT 1

AAV13493
ID AAY13493 standard; Protein; 1751 AA.

XX AAY13493;

XX 30-JUL-1999 (first entry)

XX Truncated cellulases comprising amino acid sequence.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
XX cotton-containing fabric; stonewashing.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-0810919.

XX 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;

XX Gibbs MD, Morgan H, Williams DP;

XX WPI; 1999-315403/27.

XX N-PSDB; AAX55662.

XX New truncated cellulase proteins, useful in detergents and for
XX producing 'stonewashed' denim

XX PS Claim 7; Page 37-41; 65pp; English.
 CC The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkalophilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AY13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AY13493; Cel E3/B5 is shown in AY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.
 XX SQ Sequence 1751 AA;

Query Match 6.5%; Score 334; DB 20; Length 1751;
 Best Local Similarity 21.2%; Pred. No. 2e-09;
 Matches 219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;

QY 18 SLLAATASFAVAALGLVPLTAITASPAHAA-----TTQP---YTWSNVAI 59
 Db 186 svvaet-----aslavasvvikernsqaasyqlhakdlfefadttrsdagyaatgyf 240
 QY 60 GGGGFVDGIVFNEGAPGLLYKRTD-----IGGMYRW-----DAANGRWIP 99
 Db 241 tsggfiddlgw---aawlylatndssyltkaeelmseyangntwtqcdwdrvrygltim 297
 QY 100 LLDVWGNNGYNGVSIADPINTNKNVAAVGM-YTNSWDNDGAILRSSDQG----- 152
 Db 298 lakitgkel-ykgavernidhwdritytpkmgayitgw---gslyrattaafacv 351
 QY 153 ATWQ-----ITPLPKLGNMPGR---GMGERLAVDPNNNDILYFGAPSG 194
 Db 352 adwsgcdsnkktkylnfaksgidyalgst-grsfvvgfngypqphphrn----- 400
 QY 195 KGLWRSTDGATNSQMTNFPD-----VGTIANPTDTTGYQSDIQGVVW---VAFDKSSS 246
 Db 401 -----ahsswansmkipeyhrhilygalvgpggsddsyndditdyvqnevacydnag 452
 QY 247 SLGQASTIFGVADPNPNVFWSDGGATQWQVPGAPTGFIPHKGVFDPVNVHLYIAT-- 304
 Db 453 ivgalak-----mvglygggepid--dfkaletptndeliveskf 489
 QY 305 SNTGGPYDSSGDVWKFVSTSGTWTRISVPVSTDTAN-DYF-----CYSGLTIDRQ 354
 Db 490 gnsqgp---nytevislynrwtg----prrvtdklsfkfidlteliqagys----- 535
 QY 355 HPNTIMVATQISWMPDTRIIFRSTDGGATWRIWDWTSYPNRSLRYVL-DISAPFWLTFGV 413
 Db 536 -pdvkvkvt-----yyiegkkgisgpyvwd---karniyyvlvdfsgtk-----i 575
 QY 414 QPNPVPSPKLGW-----MDEAMADIPNSDRMLYGTGATLYATNDITKWDGGQIHIAPM 469
 Db 576 ypggevehkkaqkqkiasvpgypwdptn-dpsykgltsleknkyaaydnna-----l 628
 QY 470 VKGLEETAVNDLISP-----PSGAPLISALGDLGFTHADVTAVPSTIFT---SPVFT 519
 Db 629 vwglepaaatstaptstptptptptvta-----tptptptptgsp--g 674
 QY 520 TGTSDVDAELNPSIIVRAGSFDSPSSQPNDRHVAFTDGGKNWFGSPGGVTTGTTVAAS 579
 Db 675 tsggvklyknnetsastgsirp-----wfk-----ivngg---sss 708

QY 580 ADGSRF---VM--APGDPGPVYVAVGVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNG 634
 Db 709 vdlsvrkirwytyvdgkpsavc-----dw-----adigasvnt-fnfvklsng 752
 QY 635 TFRSTDTGGVTFOQVAAAGLSSGAVGMFHAVPGKE-GDLWLAASSGLYHSTNGSSWSA 693
 Db 753 v-----sgadyy-levgf-ssgagqlq-----pgkdtgdiqvirfnkndwsnyngaddsw 800
 QY 694 ITGVSSAVNVGFKSAPSSYPVAVFVVTGGVTGAYRSDDCGTTWVLINDDQHOYGNWG 753
 Db 801 lqsmtn-----ydenakvtly-----vdg-----vlv-----wg 824
 QY 754 QAITGDHANLRVYIGTNGRGIYVDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPS 813
 Db 825 q-----epgga-----tpapstatpt 841
 QY 814 PSFSSSPSSP 873
 Db 842 ptptatpt 887
 QY 874 KVOYKNDSPAGDNOIKPGLQVNTGSSVDLSVTVTYVWFTPDGSGSTLVYNCDMAAIG 933
 Db 888 kwyangnlspstnvpkikienvttaavlsvrvywytyldgeatqsv----- 938
 QY 934 CGNIRASFGSVNPA 947
 Db 939 -----svassinpa 947

RESULT 2
 AARL5625
 ID AARL5625 standard; Protein; 782 AA.
 XX AARL5625;
 AC AARL5625;
 XX 17-MAR-1992 (first entry)
 DT Cellulase AE-1.
 DE Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
 XX Aeromonas strain no. 212.
 OS JP03251174-A.
 PN 08-NOV-1991.
 XX 28-FEB-1990; 90JP-0045465.
 XX 28-FEB-1990; 90JP-0045465.
 PA (OJIP) OJI PAPER KK.
 DR WPI; 1991-373412/51.
 DR N-PSDB; AAQ151178.
 XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 of opt. pH when carboxymethylcellulose is used as substrate.
 PS Claim 2; Fig 3; 8pp; Japanese.
 XX The sequence was deduced from the gene which was sequenced from
 CC plasmid, pAEC 1, prep. by ligating chromosomal DNA contg. the
 CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
 CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 CC methylcellulose is the substrate. The N-terminal sequence: GHADT-
 CC has been confirmed by Edman degradation. The gene can be used to
 CC produce recombinant enzyme which is used for the effective utilis-
 CC ation of biomass resources and the mfr. of pharmaceuticals and
 CC foodstuffs, and also for the detergent and deinking of waste paper.
 XX Sequence 782 AA;

XX	AC	AAB72314;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
----	----	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

XX	AAB72314;	
AC		
XX	16-MAY-2001 (first entry)	
DT		
XX	Glutamic acid rich protein-like protein amino acid sequence.	
DE		
XX	Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;	
KW	NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;	
KW	pathogen resistance; abscisic acid metabolism.	
XX	Synechocystis sp.	
OS		
XX	WO200112801-A2.	
PN		
XX	22-FEB-2001.	
PD		
XX	17-AUG-2000; 2000WO-US22961.	
XX		
PF		
XX	18-AUG-1999; 99US-0149656.	
PR		
XX	23-MAY-2000; 2000US-0206405.	
PR		
XX	(PION-) PIONEER HI-BRED INT INC.	
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Bidney DL, Crasta OR, Hu X, Lu G;	
XX		
DR	WPI; 2001-211215/21.	
XX		
XX	Novel isolated defence-related signalling gene isolated from sunflower	
PT	encoding neoxanthin cleavage enzyme, amino acid permease or glutamic	
PT	acid-rich protein useful for increasing resistance of plant to a	
PT	pathogen	
XX		
PS	Example 1; Page 128-135; 135pp; English.	
XX		
CC	This invention relates to defence-related signalling genes isolated from	
CC	the sunflower (Helianthus annuus). The genes encode a neoxanthin cleavage	
CC	enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich	
CC	protein (GRP). The signalling gene is useful for increasing the	
CC	resistance of a plant to a pathogen such as fungus, virus, bacterium,	
CC	nematode or insect (e.g. European corn borer), preferably	
CC	Sclerotinia spp., Phoma spp., or Phomopsis spp, by stably incorporating a	
CC	construct containing the gene into the genome of the plant. The gene is	
CC	useful for regulating gene expression in a plant, in response to a	
CC	stimulus such as infection with a pathogen, damage from a pathogen,	
CC	hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,	
CC	oxalic acid or expression of a gene encoding oxalic acid oxidase. The	
CC	genes are also useful for stem-preferred regulation of gene expression in	
CC	a plant. The genes are useful in agriculture, particularly in the	
CC	breeding of crop plants with improved agronomic traits, for modifying	
CC	abscisic acid (ABA) metabolism and for modifying amino acid transport and	
CC	content in plants. The present sequence represents a glutamic acid rich	
CC	protein (GRP)-like protein isolated from Synechocystis sp. The protein is	
CC	used in the characterisation of sunflower AAP.	
XX	Sequence 1749 AA;	
SQ		
Query Match 5.9%; Score 301.5; DB 22; Length 1749;		
Best Local Similarity 19.0%; Pred. No. 1.1e-07;		
Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;		
QY	17	VSLAATASFVAAALGVLPATASPAHAATTQPTWNSVAIGGG----- 62
Dd	23	lallacssfsfg---nvlaqnitpadgtgttdvdaqgnqfnigsgslgdggnlthslq 78
QY	63	-----GFVDGIVFNEGAPGILYVRTDIGMVRWDA 92
Dd	79	qfgldggqianflsnpdnrltrivggdasiingliqvsngnanflmnpagmifgpn 138
QY	93	A-----NGRWIPLLDWVGWNNWGVVVS----- 116

Db 139 sinvpdgvvttsaigfndqfvgv-----fsdndynalignpsqfardlanpplina 193
 QY 117 -----IAADPINTKNVAAVGYTNSWDPNDGAILRSS----- 149
 Db 194 gdlsvtegnlftlagnivntgslaapgnitvaavpgqnririsaqgallsleverspq 253
 QY 150 -DQATWQITPLPKFGKNNPGRMGRLAVDPND-----NIFYFGAPSGKGLWRSTD 202
 Db 254 mngggsfvlldplltltqgsnldlg-lavqngsvtngtnalvslpgsvltisgnvd 311
 QY 203 SGATWQMTNFPDVGYYIANPTTTCYSQ---DIQ-----GVVWAFD-KSSSLCQASK 253
 Db 312 as---gkstniss-ggqvalagqiaqvqatvdvsgnggggtvrigdggqitlpnasq 367
 QY 254 TIFGVGADPNPV-----FWSRDCGA--TQAVPCAPTGFPHKGVFDPVNHVLIATSNT 307
 Db 368 tli-----dsnvvkadalltnggtvivwaddstrfsgni-----seq 406
 QY 308 GGPYDSSGDVWKFVSCTWTRISVPSTDTANDYFGSLTIDRHPNTIMVATQISW 367
 Db 407 ggumgngg----fvetsg-----akslmvdtdtarvntfatmgelgt 444
 QY 368 W---POTIIFRSDDGATWTRWD-----WTSYPNRSRLRVLDISAEPLWTFGVQPNPPVP 420
 Db 445 wlldpleilvgttdlladpklvsvltittsdngnvlqadqdsiaqvafnsadpsap-- 502
 QY 421 SPKLGMDEAMAIIDPFNSRMLYGTGATLYA-----TNDLTKWDSGGQIHI-- 466
 Db 503 -gntfdspitidalfs---lgtssilfantgtpintgntlvtsptfndfankqlna 557
 QY 467 -----APWVKLEETAVN---DLISPPS-----GAPLISALGDLG----- 498
 Db 558 ntftagpydiylfrksvnggfdllgnanfvfdggagittplksfgvtateiyvgndivt 617
 QY 499 -----GPTHADVT-AVPSITFTSPVFTTG-----TSVDVYAEPLNP 531
 Db 618 gnnqifdgvyfqlpvnlltsagsvltfnillngslqvtaqniavqpsssisaveias 677
 QY 532 SIIVRAGSDPPSQPNDRHFAF---STDGKKNWFQSEPGVTTG-----GTVA 577
 Db 678 dvlilnag-----qnvsgfngintrgnvdiqal--gnistgsivtspfgnagnvi 725
 QY 578 ASADG---SRFWAPGDPQPVVAVFGNSWA-----ASQGV 612
 Db 726 lnaggtlttgyletsngtdvttssgntstayidtrfgdgleidslggavsiesskd 785
 QY 613 PANAQIRSDRVNPKTFYALSNFTYRSTDTGVTFQPVAAAGLPSSGAVGVMEHAVPCKE-- 670
 Db 786 ittafidtaysiesfnegtgngnvltagdsiltntnyiftagkngg--diffdagesiei 843
 QY 671 -----GDLWLA--SSGLYHSTNGGS-----SWSAITGV 697
 Db 844 dyIntygsqtsqgyveapldisgsviytgggpgnvnflqaggdittsyidtsaang 903
 QY 698 SSANVVG-----FGKSAPGSSYPFAVFWGT 722
 Db 904 difiqsgdtevgylftkgyegrgdvyvetgryfraidgfillgeegpfsvtagitvg- 962
 QY 723 IGGVTGAYRSDGCTWVLINDDOHOYGNWQAITGDHANLRVYICTNGR--GIYVGD 780
 Db 963 -----gsvyi-----qfggsepflignpi-----tngtgaissgd 994
 QY 781 GGAPSGSP-----SPSVSPSASPFLSPSPSSPSPPSSPSPPSSPSPPSSPS 825
 Db 995 ntvptgtpifdftldnittepepepepepepepepepepepepepepepepepepe 1054
 QY 826 PSP 868
 Db 1055 pe 1097

RESULT 4
 AAY13492

AA13492 standard; Protein; 1426 AA.
 AAY13492;
 30-JUL-1999 (first entry)
 Truncated cellulases Cel B4/5 and Cel B5.
 Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 cotton-containing fabric; stonewashing.
 Unidentified.
 EP921188-A2.
 09-JUN-1999.
 15-SEP-1998; 98EP-0810919.
 19-SEP-1997; 97US-0932571.
 (CLRN) CLARIANT FINANCE BVI LTD.
 Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 Gibbs MD, Morgan H, Williams DP;
 WPI: 1999-315403/27.
 N-PSDB; AAX55661.
 New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim
 Claim 7; Page 34-37; 65pp; English.
 The invention relates to a recombinant cellulase active protein free of
 proteinases of native thermophilic and alkaliphilic origin, comprising
 the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 length sequences, or functional equivalents, Cel B5 extends from amino
 acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 K1751 and the stability region extends from amino acid E482 to G635 in
 the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 enzymes are useful in laundry detergent compositions to prevent or
 remove staining, backstaining or graying, for use on cellulosic
 materials including cotton-containing fabrics. They are especially useful
 for preventing redeposition of colorant during stonewashing, and for
 processing of textiles where cellulose breakdown is required. The new
 truncated enzymes show reduced redeposition of dye compared to using
 non-truncated cellulase compositions.
 Sequence 1426 AA;
 Query Match 5.4%; Score 276; DB 20; Length 1426;
 Best Local Similarity 42.3%; Pred. No. 1.8e-06;
 Matches 52; Conservative 29; Mismatches 40; Indels 2; Gaps 2;
 QY 835 SRSP 893
 Db 375 saspptvtatptptvtvtatptptgtptgtgsklvlyknnetsasgsirpwf 434
 QY 894 QVVNTGSSVDLSTVTVRYWFTTRDGSSTLVVNCDAATCGGNIRASFGSVNPATPAT 953
 Db 435 kinvngssvdlrsrvkirywtvtdgdkpsav-cdwaqigasvntfnfklsgsvgady 493
 QY 954 YLQ 956
 Db 494 yle 496


```

RESULT 5
ABG27250
XX ID . ABG27250 standard; Protein; 406 AA.
XX AC ABG27250;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27241.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS91437.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX PS Claim 20; SEQ ID No 57609; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 406 AA;

Query Match 5.3%; Score 274; DB 22; Length 406;
Best Local Similarity 63.6%; Pred. No. 5.8e-07;
Matches 56; Conservative 2; Mismatches 28; Indels 2; Gaps 1;

QY 784 PSGSPSPSVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPAS 843
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 ppppspslplpplspppspspppspppspppspppspppspppspppspppsppps 156
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 844 PPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 869

```

```

Db 157 pppspslpppspppspppspppsppspi 184
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 6
AAW15238
ID AAW15238 standard; protein; 531 AA.
XX AC AAW15238;
XX DT 28-JAN-1998 (first entry)
XX DE Scaffoldin protein from Clostridium thermocellum.
XX KW Dockerin; Celd; Cipa; scaffoldin; cellulose binding domain;
XX KW chromatographic separation; soluble substrate modification; CBD;
XX KW multi-enzyme delivery system; animal feed; paper production;
XX KW plant protection; pest control.
XX OS Clostridium thermocellum.
XX FH Key Location/Qualifiers
XX FT Region 1..153 /label= internal_repeat_element_1
XX FT Region 154..306 /label= internal_repeat_element_2
XX FT Domain 239..531 /label= cellulose_binding
XX FT /note= "only 60% of the CBD is present"
XX PN WO9714789-A2.
XX PD 24-APR-1997.
XX PF 16-OCT-1996; 96WO-US16485.
XX PR 17-NOV-1995; 95US-0559968.
XX PR 17-OCT-1995; 95US-0005701.
XX PA (GEMV ) GENENCOR INT INC.
XX PI Bott RR, Clarkson KA, Fowler T, Liu C, Ward M, Xia H;
XX WPI; 1997-245106/22.
XX DR Composition with enzymes non-covalently bound to a peptide backbone
XX PT - used as a multi-enzyme delivery system, e.g. in food processing,
XX PT textiles and pest control
XX PS Claim 10; Fig 6; 29pp; English.
XX CC This protein sequence represents a scaffoldin based on the Cipa protein
XX CC of Clostridium thermocellum. The scaffoldin protein is used in a novel
XX CC composition that comprises at least 2 enzymes non-covalently bound to a
XX CC peptide backbone (i.e. present sequence). The scaffoldin comprises a
XX CC number of internal repeating units and at least one cellulose binding
XX CC domain (CBD). The CBD may be altered to modify its affinity for
XX CC cellulose, which may be desirable where cellulose binding would be
XX CC disadvantageous. The enzyme is bound to the scaffoldin by a dockerin
XX CC region of the enzyme, which binds to a repeating element of the
XX CC scaffoldin. The dockerin is preferably Cels (AAW15237) or Celd
XX CC (AAW15236). The composition can be used in reducing allergenicity,
XX CC producing synergistic effects, and facilitating selective modification
XX CC of substrate. By taking advantage of the cellulose binding domain of the
XX CC complex, the complex could be immobilised for use in chromatographic
XX CC separations or for soluble substrate modification. By adding the
XX CC scaffolding domain, it is possible to recover enzymes, or to quantify
XX CC the amount of an enzyme in a solution. The composition could also be
XX CC used in a multi-enzyme delivery system which could be used in the food
XX CC industry, in food processing, animal feed, textiles, bioconversion,
XX CC pulp and paper production, plant protection and pest control, as a wood
XX CC preservative, topical lotions, and biomass conversions.

```


SQ	Sequence	531 AA;	
DR	N-PSDB; AAT86625.		
XX	Cellulase proteins with cohesin or dockerin type II domains - useful for potentiating the activity of multiprotein enzyme complexes		
PT	Disclosure; Page 47; 60pp; French.		
XX			
XX	Multimeric protein, especially enzymatic, complexes are held together by protein-protein interactions between domains designated dockerins and cohesins, which are found on the catalytic and scaffold subunits respectively. An example of such a complex is the cellulose degrading protein complex from Clostridium thermocellum, known as the cellulosome. This complex comprises around 15 proteins including endoglucanases, cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which interact with a central "scaffold" protein designated the cellulosome integrating protein (Cip; see AAW43108). The catalytic subunits interact with the Cip subunit via conserved 23 amino acid dockerin domains. Cip has been shown to contain 9 copies of a cohesin domain. The invention relates to the isolation of proteins binding to a novel dockerin type domain found in the C-terminal portion of Cip. The new domain is designated a type II dockerin domain (as compared to the type I domain found on the catalytic subunits of the cellulosome). The type II dockerin domain has some sequence similarity to the type I dockerins but is unable to bind type I cohesin domains.		
CC	The sequence presented here is the cellulosome integrating protein (CIP-A) to which the proteins SdbA, OipB and ORF2p (AAW43105-W43107) can bind at the C-terminal dockerin type II domain (location not given in specification). The novel type II dockerin and cohesin domains can be used in complexes, especially enzyme complexes, to potentiate their catalytic actions in a synergistic manner.		
XX	Sequence	1853 AA;	
SQ			
Query Match 5.1%; Score 261.5; DB 19; Length 1853;			
Best Local Similarity 23.2%; Pred. No. 1.4e-05;			
Matches 117; Conservative 74; Mismatches 159; Indels 155; Gaps 20;			
QY	508 VPSTFTSPVFTTG---TSVDYAEALNPSTIIVR-----AGSFDPSQPNDRHVA--FSTDGG	558	
DB	2 VPSKGMANCDVIGYDPNVLEVTEVKPGSIKDPDPSKFSDSAIPDKRMIVLFAEDSG	61	
QY	559 KNMFQSGEPGGVTTGGTVAASADGRFVWAPDGPQPVVYAVFGNSWAASQGVPAANAQI	618	
DB	62 I r-----gtyaitqdg-----vfat-----ivaty	80	
QY	619 RSDRVNPKTFYALNNGTFYRS-----TDGGV-----TFQPVAAAGLPSSGAV---	659	
DB	81 KSAAPITL--LEVGAFAENDLVEISTTFVAGGVNLGSSVPTTQP---NVPDGVVVEI	135	
QY	660 -----GVMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG	706	
DB	136 gktgsvgttveipvyfrgvpks-----glancdfvrydpnvleigdigpddiivdpnpt	191	
QY	707 KSPAGSSYP---AVFVGTIGGVTGAYRSDDCGTTWLINDQHQYGNWGOAITGD--H	760	
DB	192 KSFDTALYPRKLIIVLFAEDSG-TGAY-----aItKdgvf	226	
QY	761 ANLRRVYIGTNGRGIYVGDIGGAPSGSPSVSPSASPSPSPSPSPSPSPSPSPSSP	820	
DB	227 akiratvksaapgyitfdevggfad-----ndlveqkvsvfiddggnvgnatcptkgatp	279	
QY	821 SSPPSP	880	
DB	280 tnatptksatapttr-----psvptntptantpantpsvgnlkvefyns	323	
QY	881 DSAPGDNQIKPGLQVYVNTGSSVDLSTVTVRYWFTTRDGGSTLVYNCDAIAI-----	932	
DB	324 npsdttnsinpgfkvtntgssaidlskltrlyyytvdgqkdqtfw-cdhaaigsnsgsyn	382	
QY	933 -GCGNIRASFGSVNPTPTADTYLQ	956	
DB	383 gitsnvkgtfvkmsstnnadtyle	407	
RESULT	7		
AAW43108			
ID	AAW43108 standard; Protein; 1853 AA.		
XX			
AC	AAW43108;		
XX			
DT	16-OCT-1998 (first entry)		
XX			
DE	C. thermocellum cellulosome integrating protein.		
XX			
KW	Multimer; enzyme; complex; protein-protein interaction; dockerin domain;		
KW	cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;		
KW	cellulosome integrating protein; scaffoldin dockerin binding protein.		
XX			
OS	Clostridium thermocellum.		
XX			
PN	FR2748479-A1.		
XX			
PD	14-NOV-1997.		
XX			
PF	10-MAY-1996; 96FR-0005854.		
XX			
XX	10-MAY-1996; 96FR-0005854.		
PR	(INSP) INST PASTEUR.		
XX			
PA	Beguin P, Leibovitz E;		
XX			
PI	WPI; 1998-011569/02.		
XX			
DB			

RESULT 8

AAW43106
ID AAW43106 standard; Protein; 1664 AA.

XX AC AAW43106;

XX 16-OCT-1998 (first entry)

XX DE C. thermocellum OlpB protein.

XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW cohesin domain; catalytic subunit; scaffold subunit; SAbA; synergistic;
KW cellulosome integrating protein; scaffoldin dockerin binding protein.

XX Clostridium thermocellum.

XX Key Location/Qualifiers

FT Domain 28..192 /note= "cohesin type II domain"

FT Domain 207..363 /note= "cohesin type II domain"

FT Domain 409..565 /note= "cohesin type II domain"

FT Domain 607..763 /note= "cohesin type II domain"

FT Domain /note= "cohesin type II domain"

XX FR2748479-Al.
PN 14-NOV-1997.

XX PD

XX PF

XX 10-MAY-1996; 96FR-0005854.

XX 10-MAY-1996; 96FR-0005854.

XX (INSP) INST PASTEUR.

XX Beguin P, Leibovitz E;

XX WPI; 1998-011569/02.

XX N-PSDB; AAT86623.

XX Cellulase proteins with cohesin or dockerin type II domains - useful
PT for potentiating the activity of multiprotein enzyme complexes
PS Claim 7; Page 31-39; 60pp; French.XX Multimeric protein, especially enzymatic, complexes are held together
CC by protein-protein interactions between domains designated dockerins
CC and cohesins, which are found on the catalytic and scaffold subunits
CC respectively. An example of such a complex is the cellulose degrading
CC protein complex from Clostridium thermocellum, known as the cellulosome.
CC This complex comprises around 15 proteins including endoglucanases,
CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
CC interact with a central "scaffold" protein designated the cellulosome
CC integrating protein (CipA; see AAW43108). The catalytic subunits
CC interact with the Cip subunit via conserved 23 amino acid dockerin
CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
CC The invention relates to the isolation of proteins binding to a novel
CC dockerin type domain found in the C-terminal portion of Cip. The new
CC domain is designated a type II dockerin domain (as compared to the type
CC I domain found on the catalytic subunits of the cellulosome). The type
CC II dockerin domain has some sequence similarity to the type I dockerins
CC but is unable to bind type I cohesin domains.XX The sequence presented here is an example of a protein which binds
CC the novel type II dockerin domain and is the product of the OlpB gene.
CC The protein contains 4 type II cohesin domains in the N-terminal portion
CC of which the first domain (amino acid residues 28-192) is thought to
CC bind CipA. The novel type II dockerin and cohesin domains can be used
CC in complexes, especially enzyme complexes, to potentiate their catalytic
CC actions in a synergistic manner.

XX

SQ Sequence 1664 AA;

Query Match 4.8%; Score 244; DB 19; Length 1664;

Best Local Similarity 20.8%; Pred. No. 0.00011;

Matches 185; Conservative 96; Mismatches 316; Indels 292; Gaps 44;

QY 100 LLDWGVNNWGYN---GVVSTAAADPINTNKVAAVGMVNTNSWDPNNDGAILRSSDGGATW 155

DB 174 lfdwdaetitgeviqpdliivveaeplkd-----asvaleldktkvkgvdiitat----- 223

QY 156 QITPLPFKLGNNMPGGMGERLAVDPNNDNI--LYFGAPSGKGLWRSTDSGATWSQMTNF 213

DB 224 -----ikienmknfagynlnikydpmtlealeletgsaiaakrtwpyt--ggvtlsdsny 275

QY 214 PDVGTYIANTPTDTGYOSDIQGVVWVAFDKSSSSLSQASAKTIFGVADPNPNV-----FWS 269

DB 276 -----gkttavandv-gagiinfaeaysnlkyret---gvaeeetgiigkigfrv 321

QY 270 RDGGAT-----WQAVPGAPTGFIPHKGVDPVNHVLYIATSTNGTGGPYDGSDDGVKFSV 323

DB 322 lkagstairfedttampgaletgy-----mfdwyg-----enikgysvvqgeivaeg 370

QY 324 TSGTWTIRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQI-SWMPDTIIFRSTGGAT 382

DB 371 epgeeteepvp-tetpvd-----ptctvteepvpseipdsyvimeld--kt 414

QY 383 WTRIDWTSYV-----NRSRYVLDISAEPWLTFGVQPNPPVPPSKLGMDEAMAIDPF 436

DB 415 kvkvgdiiatikiemknfagynlnikydpmtlealeletgsaiaakrtw-----pv 466

QY 437 NSDRMLYGT--GATLYATNDLTKWDSGGQIHIAPMVKGL---BETAVNDLISPPSGAPLI 491

DB 467 tggvtlqsdnygkttavandv-----gagiinfaeaysnlkyretgv-----aeet 513

QY 492 SALGDLG-----GFTHADVTAVPSTIFTSPVT-TGTSV-DYAEALNPSIIVRAGSF 540

DB 514 gliigkigfrvlkagstairfedttampgaletgyfwdygenikgysvvqgeivaeg- 572

QY 541 DPSSQP-----NDRHVAFTDGGKWNFGSEPGGVTTGGTVAA 578

DB 573 epteeptetpvdptptvteepvpseipdsyvimeldkt-----vkegdvila 622

QY 579 SADGSRFVWAPDGPQPVVYAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALS----- 632

DB 623 tirvnniknlag-----yqigi-----kydpkvlfaetnietgdp 656

QY 633 -----NGTFYRSTD---GGVTFQPVAAAGLPSSGAVGVMF--HAVPGKEGDLWLAAS 678

DB 657 idegtwpavgggtllknrdylptgvainnvskgilnfaayvyvfdyreeeksd----- 710

QY 679 SGLYHSTNGSSWSAITGVSSAVNVGFG-----KSAPGSSYPVAVVVGITG 724

DB 711 -----tglig--nigfrvlkaedttirfeeslempgs----- 740

QY 725 GVTGAYRSDDCGTTWVLIINDDQHQYCNWGOATGCDHANLRVYIGTNGRGIYVGTGGAP 784

DB 741 -idgtymld-----wyl-----nrlsg-----yvvigp 764

QY 785 ----SGSPSPSVSPASPSLSLSPSPSSSPSPS--PSPSSPSSSPSP-SPSPSPSPSPSR 837

DB 765 ikaasdepiptdpsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsde 824

QY 838 PSPASPS 886

DB 825 ptpsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsdeptsde 872

RESULT 9

ABB71513

ID ABB71513 standard; Protein; 950 AA.

XX ABB71513;

AC ABB71513;

QY 783 APGSPSPSVPSASPSLSPSPSSSPSPSPSSSSSPSPSPSPSPSRSPSPA 842

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 39.87 Seconds
(without alignments)
586.288 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPATPTADTYLOX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%
Maximum Watch 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES										
Result No.	Score	Query Match %	Length	DB ID	Description					
1	334	6.5	1751	4	US-09-136-574A-44	Sequence 44, Appl				
2	276	5.4	1426	4	US-09-136-574A-43	Sequence 43, Appl				
3	225.5	4.4	551	2	US-09-033-537A-1	Sequence 1, Appli				
4	210.5	4.1	521	1	US-08-276-213-3	Sequence 3, Appli				
5	208.5	4.1	493	4	US-09-198-956-10	Sequence 10, Appl				
6	208.5	4.1	493	4	US-09-198-955A-12	Sequence 12, Appl				
7	200.5	3.9	700	2	US-07-862-588B-2	Sequence 2, Appli				
8	199.5	3.9	490	4	US-09-109-841-2	Sequence 2, Appli				
9	199.5	3.9	616	4	US-09-136-574A-47	Sequence 47, Appl				
10	191.5	3.7	412	1	US-08-313-288B-18	Sequence 18, Appl				
11	189	3.7	933	3	US-08-293-728-2	Sequence 2, Appli				
12	189	3.7	933	3	US-09-421-868-2	Sequence 2, Appli				
13	186	3.6	423	2	US-08-760-797A-1	Sequence 1, Appli				
14	185.5	3.6	424	4	US-08-932-929B-1	Sequence 1, Appli				
15	185.5	3.6	1719	2	US-08-459-568-4	Sequence 4, Appli				
16	185.5	3.6	1719	2	US-08-399-411-4	Sequence 4, Appli				
17	185.5	3.6	1719	3	US-08-516-859A-4	Sequence 4, Appli				
18	185	3.6	334	6	5202236-3	Sequence 4, Appli				
19	184	3.6	331	6	5202236-37	Patent No. 5202236				
20	181	3.5	167	5	PCT-US95-13813-9	Sequence 9, Appli				
21	180.5	3.5	424	2	US-08-760-797A-3	Sequence 3, Appli				
22	180.5	3.5	424	4	US-08-932-929B-3	Sequence 3, Appli				
23	179.5	3.5	1481	2	US-08-616-844-40	Sequence 40, Appl				
24	179.5	3.5	1481	2	US-08-599-654-40	Sequence 40, Appl				
25	179.5	3.5	1481	3	US-08-944-868A-40	Sequence 40, Appl				
26	179.5	3.5	1481	3	US-08-944-423A-40	Sequence 40, Appl				
27	179.5	3.5	1481	3	US-08-944-496-40	Sequence 40, Appl				

28	178	3.5	214	1	US-08-217-327-4	Sequence 4, Appli
29	176.5	3.4	254	1	US-07-667-276A-6	Sequence 6, Appli
30	175	3.4	126	1	US-08-395-602A-4	Sequence 4, Appli
31	175	3.4	126	2	US-08-021-625D-4	Sequence 4, Appli
32	175	3.4	531	2	US-07-862-588B-7	Sequence 7, Appli
33	175	3.4	1581	4	US-09-110-517-2	Sequence 2, Appli
34	171	3.3	303	4	US-08-818-112-92	Sequence 92, Appl
35	171	3.3	303	4	US-08-818-111-93	Sequence 92, Appl
36	171	3.3	303	4	US-09-056-556-92	Sequence 92, Appl
37	171	3.3	786	4	US-09-103-429A-3	Sequence 3, Appli
38	168.5	3.3	162	3	US-08-983-045-2	Sequence 2, Appli
39	168.5	3.3	560	3	US-08-983-045-4	Sequence 4, Appli
40	168.5	3.3	772	2	US-08-410-784A-5	Sequence 5, Appli
41	167	3.3	160	6	5219987-7	Patent No. 5219987
42	167	3.3	163	5	PCT-US93-08435-2	Sequence 2, Appli
43	167	3.3	164	5	PCT-US93-08435-1	Sequence 1, Appli
44	167	3.3	1064	1	US-08-642-255-62	Sequence 62, Appl
45	166.5	3.2	805	4	US-09-103-429A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Db 435 KIVGGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQIGASNTFNFVKLSSGYSYADY 493
Qy 954 YLQ 956


```

Db 41 LWSRDYRSLDQIKS--LGYNTIRLPYSDDILKPGTMPNSINFYQMNDLQGLTSLOQVMD 98
QY 481 LIPSPGAPLISALDGLG-----GFHADVTAVPSIFTSPTVTTGTSVDYAEI----- 529
Db 99 KI-----VAGQIGLRIILDRHRPCGSGSALWYTSVSEATWISDLQALQARYKG 150
QY 530 NPSIIRAGSFDPSSOPNDRHVAFSDGGKNWFQGGSEPGG-----VTTGTVAA 578
Db 151 NPTVV---GFDLHNEPHDPACWCGDPSIDWRLAERAGNAVLSVNPNDLIFVEG--VQ 204
QY 579 SAGSRVRVAPGDPQPVVYAVGEGNSAASQ-----GVP-----ANAQIRSDRVNPKTFY 629
Db 205 SYNGDSYMWG-----GNLQAGAGYPPVVLNPNRVLVYSAHDYATSVYPTQWTF 250
QY 630 ALSNGFYRSTGGVTGTFQPVAAAGLPSGAGVGMFHAVPCKEGDLWLAASGLYHSTNGGS 689
Db 251 --SDPFF-----PNN-----MPGIWNKNW-----GYLFNQNIAP 277
QY 690 SWSAITG--VSSAVNVGFGKSAPGSSYPVAVVGTIGGVTGAYRSD-----DCG 736
Db 278 VWLGERGTTLQSTTDQTLWKLTVQYLRP-----TAQYGADSFQWTFWSWNPDSG 326
QY 737 TTWVLNDDHQYGNNGQAITGDHANLRVYICTNGRGIVYGDIGGAPSGSPSPSPSA 796
Db 327 DTGGILKDD-----WQTVDTVRKDGYLAPT-----KSSIFDPVGA-----SA 362
QY 797 SPISLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 856
Db 363 SP--SQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 418
QY 857 SSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 915
Db 419 ASGARCTAS-----YQVNSDWNGGFT-----VTVAVNTSGSVATKTTWTSWTF- 461
QY 916 RDGSSSTLVYNCDAWAA 931
Db 462 --GNGNOTITNS--WNA 473

```

```

RESULT 5
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; EARLIER FILING DATE: 1998-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 4.1%; Score 208.5; DB 4; Length 493;
Best Local Similarity 23.5%; Pred. No. 3.6e-06;
Matches 103; Conservative 62; Mismatches 149; Indels 125; Gaps 19;

QY 568 GGYTTGGTVAASAD-----GSRFWAPDGPQPVVYAVGFG-NSWAASQGVPAANAQIRSDR 622

```

```

Db 63 GTTTSNTSASKIDVXKDVSNVSVGSGTKE-----LKGIGIKIWRANNIIRNLKIHE-- 116
QY 623 VNPRTFYALNSGTFFYSTDGGVTTFQPVAAAGLPSGAGVGMFHAVPCKEGDLWLAASGLY 682
Db 117 -----VAGQIGLRIILDRHRPCGSGSALWYTSVSEATWISDLQALQARYKG 142
QY 683 HSTNGGSSW-----SAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGVTGAVR 731
Db 143 HSLNVKDYDGLFDVYKRDAAEYITFSWNVYVHDGWSMLMGSS-----D 185
QY 732 SDCGTTWVLND-----DOHGYGNNGQAITGDHANLR---RVYIGTNGR 773
Db 186 SDNYNRTITFHHWFFENLNSRVPFSFRFEGEHIYNNFNKIIDSGINSRMCARIENN-- 243
QY 774 GIYVGDIGAPSGSPSPSVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSP 826
Db 244 -----LFENAKDPIVSWYSSSPGYWHVSNKFNKRVNKGSMPTTSTTYNPPYSLD 294
QY 827 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 886
Db 295 NVDNVKSIVKQ---NAGVGKIQRPPPTPTSPPS---ANTPVGNLKVFEYNSNPSDTT 348
QY 887 NQIKPGLQVNTGSSVDLSTVTYRVYFTRDGSSTLVYNCDAWAI-----GCGNI 937
Db 349 NSINPQFKVTNTGSSAIDLKTLRYVYTVDGOKDQTFW-CDHAAIIGNSYNGSITSNV 407
QY 938 RASFGSVNPTATPDYLO 956
Db 408 KGTFVKNSSSTNNADTYLE 426

```

```

RESULT 6
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR FILING DATE: 1997-11-24
; PRIOR FILING DATE: 1997-11-24
; PRIOR FILING DATE: 1997-11-24
; PRIOR FILING DATE: 1997-11-24
; PRIOR FILING DATE: 1997-12-02
; PRIOR FILING DATE: 1997-12-02
; PRIOR FILING DATE: 1998-05-06
; PRIOR FILING DATE: 1998-11-02
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 4.1%; Score 208.5; DB 4; Length 493;
Best Local Similarity 23.5%; Pred. No. 3.6e-06;

```


US-09-109-841-2

[illegible]

RESULT 9
 US-09-136-574A-47
 ; Sequence 47, Application US/09136574A
 ; Patent No. 6294366
 ; GENERAL INFORMATION:
 ; APPLICANT: Farrington, Graham K.
 ; Anderson, Paige
 ; Gibbs, Moreland
 ; Bergquist, Peter
 ; Daniels, Roy
 ; Morgan, Hugh W.
 ; Williams, Diane P.
 ;
 ; TITLE OF INVENTION: Compositions and Methods for
 ; Treating Cellulose Containing
 ; Cellulase Enzyme Compositions
 ;
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Howson and Howson
 ; STREET: Spring House Corporate Center, P.O. Box 457
 ; CITY: Spring House
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19477
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/136,574A
 ; FILING DATE: 19-Aug-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/932,571
 ; FILING DATE: September 19, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bak, Mary E.
 ; REGISTRATION NUMBER: 31,215

[illegible]

```

RESULT 10
US-08-313-288B-18
; Sequence 18, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and AviHu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-313-288B-18

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 2, 2002, 09:11:52 ; Search time 54:67 Seconds
(without alignments)
1682.047 Million cell updates/sec
Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPTPTADTYLQX 957
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	
1	2436	47.4	890	2	T35237	probable secreted
2	2037	39.7	839	2	D97013	probably secreted
3	1740	33.9	856	2	T00349	Avicelase III - As
4	1161	22.6	707	2	F72393	hypothetical prote
5	431.5	8.4	1779	2	T31085	xylanase - Caldice
6	398.5	7.8	1711	2	T31337	1,4-beta-glucanase
7	397.5	7.7	1742	2	T17120	cellulase (EC 3.2.
8	372.5	7.3	473	2	S50755	hypothetical prote
9	346.5	6.7	901	2	A49227	sialidase - Actino
10	340.5	6.6	1331	2	A48954	mannan endo-1,4-be
11	308.5	6.0	611	2	S76211	hypothetical prote
12	301.5	5.9	1749	2	S75138	hypothetical prote
13	293	5.7	279	2	T10361	hypothetical prote
14	291	5.7	351	2	S50754	hypothetical prote
15	290.5	5.7	2232	2	T34434	hypothetical prote
16	290	5.6	1039	2	S02711	cellulase (EC 3.2.
17	288.5	5.6	1032	2	T34433	hypothetical prote
18	283.5	5.5	474	2	S15921	protein TPX-VT3 -
19	262	5.1	2468	2	A83412	hypothetical prote
20	261.5	5.1	1854	2	S36859	c1pA protein - Clo
21	254.5	5.0	913	2	S20590	exo-alpha-sialidas
22	246	4.8	915	2	A43802	cellulase (EC 3.2.
23	245	4.8	879	2	A47704	endoglucanase I (E
24	244	4.8	1664	2	T18262	S-layer protein -
25	243.5	4.7	552	2	T08148	proline-rich myros
26	243	4.7	288	2	T17737	proline-rich prote
27	235.5	4.6	2124	2	A28452	proteoglycan core
28	229.5	4.5	496	2	T17908	proline/lysine-ric
29	227.5	4.4	2271	2	F90073	hypothetical prote

30	222.5	4.3	360	2	S12850	protein TPX - Ther
31	222.5	4.3	3570	2	T45025	mucin MUC5B, trach
32	221	4.3	2132	1	A55182	aggreccan precursor
33	219	4.3	4776	2	E95206	cell wall surface
34	218.5	4.3	3164	1	WMBEH6	UL36 protein - hum
35	217	4.2	241	2	AC2284	hypothetical prote
36	216	4.2	383	2	T46707	proteophospholyca
37	216	4.2	449	2	S16748	proline-rich prote
38	215.5	4.2	13055	2	T16580	hypothetical prote
39	215	4.2	339	2	T17636	proline-rich prote
40	215	4.2	13288	2	T03099	mucin, submaxillar
41	214	4.2	300	2	J02220	hydroxyproline-ric
42	214	4.2	532	2	B35621	spore germination
43	211.5	4.1	1275	2	T33369	hypothetical prote
44	211	4.1	602	2	AD2067	hypothetical prote
45	210	4.1	549	2	T17525	proline-rich prote

ALIGNMENTS

RESULT 1
T35237
probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35237
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21572
A:Accession: T35237
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-890 <SEE>
A:Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.30C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5C7.30C

Query Match		47.4%	Score 2436;	DB 2;	Length 890;
Best Local Similarity		48.6%	Pred. No. 1.8e-102;		
Matches		475;	Conservative 129;	Mismatches 248;	Indels 126; Gaps 19;
QY	11	MSRRRLVLLAATASPAFAAALGVLP	IAITASPA--HAATTPQYTWNSVAIGGGFVDGI	68	
Db	1	MRTRILTLLAALAGLLA---GSPPAASAAEPAPRAA	VAADSYTWKNARIDGGGFVPGI	57	
QY	69	VNEGAPGLIYVRDIDGGMYRDAANGRWIPLD	WVGWNNWYGVVSTAADPINTKMW	128	
Db	58	VFNRTKDLAYARTDIDGGAYRWQEESTW	TPLLDHVGWDDWGHGTGVVALASDAVDPRVY	117	
QY	129	AAVGMYTNSWDNDGAILRSSDQGTWQITP	LPFKLGGNMPGRGMERLAVDPNNILY	188	
Db	118	AAVGTINDWDPTNGAVLKSADRGASWEKAD	LPFKLGGNMPGRGMERLAVDPHDNDVLY	177	
QY	189	FGAPSGKGLWRSDTSQWTFNQFDPVGT	YIANPTDTTGYQSDIQGVVWVAFDKSS--SS	247	
Db	178	LGAPSGHGLWRSDTACGVTSEVTA	FPNPGYADPNDTSGYASDNQGITWTFDESTGGG	237	
QY	248	LQASAKTIFVGADPNPNVFWSRDGGAT	QWAVPGATGTPHKGVPDPNVHVIYATSN	307	
Db	238	AGTATRTLVGVADKENAVYRSTAGAT	WERLAGQPTGYLAHKGVLDAGENGLYLA	SDT 297	
QY	308	GGPYDSSGDVWKFSTGVTWRISVP	STDTANDYFGYSGLTIDRQHPNTIMVATQISW	367	
Db	298	GGPYDGGKRLRYATATGTWTDIS	PAEADT---YIGFSGLTVDQRQCTVMATYSW	354	
QY	368	WPDITIFRSDGGATWTRIDWTD	SYPNRSLRYVLDISAPFWLTFFGVQPNPPVSPKLGWM	427	
Db	355	WPDQTFIRSTDGSATWSQAWSY	SYDPDRENYTMDVSSSPWLTWGANPAPPEQTPKLGWM	414	
QY	428	DEMAIDPFNSDRMLYGTGAT	LYATVNDLTKW--DSGGQIHIAIPWVKGLEETAVNDLISPPS	486	

Db 415 TEALEIDPFSDRMVCTGATVYGTENLTWNDDGGTFVAPMVRGLEETAANDLASPPS 474
QY 487 GAPLISALGDLGFTHADVTAVPSTIFTSPVFTGTSVDYAEALNPSLIIRAGSFDPSQP 546
Db 475 GAPLISALGDLGFTHADVTAVPSTIFTSPVFTGTSVDYAEALNPSLIIRAGSFDPSQP 532
QY 547 NDRHAFSTDDGKNVFGSGPGVTTGGTVAASADGSRFVWAPCDPQPVVYAVGFGNSW 606
Db 533 --HIAFSTDNGANWFGTDPGSGVGGTVAACADGSRFVWSP--EGAGVQVYTTGFTSW 587
QY 607 AASQGVPPANAQIRSDRVNPKFTFYALNGFTYRSTDDGGVTFQPVAA--GLPSSGAVGVNFA 665
Db 588 QASTGLPAGAIIVSDRVNPKFTFYALNGFTYRSTDDGGVTFQPVAA--GLPSSGAVGVNFA 645
QY 666 VPKEGDLWLAASS-----GLYHSTNGSGSSALITGVSSAVNVGFGKSPAGSSYPVAVFVV 720
Db 646 LPGEEDVWLAGAAGDPYGLWISTDGGTFTRLPGVDAADTVFGKAAPGASVQTLFTS 705
QY 721 GTGGVTVAYRSDCCGTTWYLLINDDQHQYGNWQQAITGDHANLRVYIGTNGRGIYVGD 780
Db 706 AEIGGVGRIPRSTADAGATVRVNDAAHQWGTCAAITGDPVYGRVYVATNGRGIYGD- 764
QY 781 GGAPSGSPSPVSPASPSLSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
Db 765 -----TSDTGGTDPGPGDPPTP----- 782
QY 841 SASP 900
Db 783 -----TGACEVYTYTNQHPGGFQ--ADVLNTNGT 811
QY 901 S-----SVDLS-----TVTVRYWFR--DGGSTLTVNCDM-----AAGCGNIRA--- 939
Db 812 SANWGSLLDMSFPGGQEVETRMNAEHTQAGTSVARNVGNAGVAPGASVGFGTGSRSG 871
QY 940 -----SFGSVNPATPTA 951
Db 872 TNAEPGFAVAGRACPTA 889

RESULT 2
D97013
probably secreted sialidase, several ASP-boxes and dockerin domain [Imported] - Clostrid
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: D97013
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, O.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97013
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:gl5023820; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;
Best Local Similarity 49.5%; Pred. No. 1.5e-84;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

QY 13 SRLVSLAA-----TASFAVAALGVLPATITASPAHAAT-TQPYTWSNVAIGGGGFVD 66
Db 2 NKRIVSWAGLSIIFTGFE-----VTHISAANKAAASVQSGYKWDNAKTAGAGGYVP 54
QY 67 GIVENEGAPCLVVRTDIGYRWDAANGRWIPLLDVGNNVNGVVSIAADPINTNK 126
Db 55 AVIFNTEKDLIYARTDMGAYRWKANKNWIPTD--GFSDWTMLGCSIAFDPIDTNR 112

QY 127 VMAAVGMYNTSNPDNDGAILRSSDQATWQITPLPKLGNMFGRMGERLAYDPNDNNI 186
Db 113 VYTAAGLYTNDMDENAYILSSQDKGTWKRYQLPKFVGNGMFGRMGERLQIDPNDKI 172
QY 187 LYFGAPSGKLRSTOSGATWSQNTFPDVGTYIANPTDTTGYQSDIQGVWVAFDKSS 246
Db 173 LYLGAASGKLRSTOSGATWSQNTFPDVGTYIANPTDTTGYQSDIQGVWVAFDKSS 230
QY 247 SLGQASKTIFVGVAD--PNNPVFMSRGGATWQAVCAPGTGFIHKGVDPDNNHVLVIAT 305
Db 231 TKGSPQTQTVFVGAADKTNIIYVNDGGKTSVSAKQPKGLPHHGL--ASDGLLISYS 289
QY 306 TTGGPPYDGGSDGDKFSTSGTWTTRISPVSTDTANDFYSGLTIDRQHPNTIMVATQI 365
Db 290 NTCGPPYDGGSDGDKFSTSGTWTTRISPVSTDTANDFYSGLTIDRQHPNTIMVATQI 346
QY 366 SWPDPPIIRSTDDGATWTRINDWTSYPNRSLRYVLDLSAEPHLTFG--VQNPVPVSPKL 424
Db 347 RWPDPPIIRSTDDGATWTRINDWTSYPNRSLRYVLDLSAEPHLTFG--VQNPVPVSPKL 406
QY 425 GWMDEAMADPFNSDRMLYCTGATLYATNDLTKWDSGGQIHIAPVMKGLFEETAANDLIS 484
Db 407 GWMGDLIDPFNSDRMLYCTGATLYATNDLTKWDSGGQIHIAPVMKGLFEETAANDLIS 466
QY 485 PSGAPLISALGDLGFTHADVTAVPSTIFTSPVFTGTSVDYAEALNPSLIIRAGSFDPS 544
Db 467 TGAQALLSAGVDDCGFYHDDITKVPKMMTTPNFSAATTSIDYAESVPNFVYRNGVDTSK 526
QY 545 QPNDRHVAFTDGGKWNFO--GSEPGGVTTGGTVAASADGSRFVWAPCDPQPVVYAVGFG 603
Db 527 NQDDKCGISYDGGKWNFSAGSNISGVYKAGTVAAGADAKTIVWSP--EGANNAISTDNG 585
QY 604 NSWAAASQGVPPANAQIRSDRVNPKFTFYALNGFTYRSTDDGGVTF--QPVAAGLPSSGAVGV 662
Db 586 NKWTPCSGLPQGAQKVRSDRVNPKFTFYALNGFTYRSTDDGGVTF--QPVAAGLPSSGAVGV 643
QY 663 FHAVPQEGDLWLA--ASSGLYHSTNGSGSSALITGVSSAVNVGFGKSPAGSSYPVAVFVV 721
Db 644 FKTIVIGHEGDIWAGKGLWHTSDGATFTKVSVDASDTVGLGSKTKDDGYPAIYMDA 703
QY 722 TTGGVTVAYRSDCCGTTWYLLINDDQHQYGNWQQAITGDHANLRVYIGTNGRGIYVGDIG 781
Db 704 TIDGTAGIFRSDDEGATWVRINDAHQYQSPDYCITGDPNKYGRVFGVNGRGIYVGDID 763
QY 782 GAPSGSPSPSPVS 793
Db 764 GS-QPTPTFSVT 774

RESULT 3
T00349
Avicelase III - Aspergillus aculeatus
C:Species: Aspergillus aculeatus
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C:Accession: T00349
R:Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
submitted to the EMBL Data Library, June 1998
A:Description: Avicelase III from Aspergillus aculeatus.
A:Reference number: Z14141
A:Accession: T00349
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-856 <ARA>
A:Cross-references: EMBL:AB015511; NID:d1199887; PID:d1029971
C:Genetics:
A:Gene: avIII
C:Superfamily: fungal cellulose-binding domain homology
F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;
Best Local Similarity 43.5%; Pred. No. 3.5e-71;
Matches 366; Conservative 139; Mismatches 297; Indels 40; Gaps 18;

[illegible]

A:Reference number: A72200; MUID:99287316
A:Accession: F72393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-707 <ARN>
A:Cross-references: GB:AE001712; GB:AE000512; NID:g4980799; PIDN:AAD35393.1; PID:g4980799
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0305

Query Match 22.6%; Score 1161; DB 2; Length 707;

[illegible]

RESULT 5
T31085
xylanase -
C;Species:
C;Date: 02

C;Accession: T31085
R;Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
submitted to the EMBL Data Library, December 1997
A;Description: Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. Rt69B.1.
A;Reference number: Z20972
A;Accession: T31085
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-1779 <MOR>
A;Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1
C;Genetics:
A;Note: xynC

Query Match	8.48;	Score 431.5;	DB 2;	Length 1779;
Best Local Similarity	19.86;	Pred. No. 5.7e-12;		
Matches 252;	Conservative 162;	Mismatches 182;	Indels 477;	Gaps 53;
Qy	65	VDGIVFNEGAPGILYV----	RTDIGMYRWDAAANGRWIPLLLDWYGNWNGYGVVIAAD	120
Db	20	VTILFHQEAAXAYTVDFEGTDILSPFAYGKSNIA-----	VDMGN-AYNGKSIRVS	71
Qy	121	PINTNKVMAAYGMYTNSWDPNDDGAILRSSDQCATWQIT-----	PLPFKL-----G	165
Db	72	--NRSSIWGVAV-----	DVKNTMNGTTWVVSAYVKHSYOKPVAFGISAVIDDG	119
Qy	166	GNMFGRGMERLAVDPNNDNLTLYFCAPSGKGLWRSTDGATWSQNTNPDVGTYIANTPD	225	
Db	120	SGVKTSLIEVVAI- PN-----	YWKKT-----VKMTPNIS	150
Qy	226	TTGYQSDIQGVVWVAFDKSSSSLSGQAKTIFGVADPN-----	NPVFWSR--DGGAT--WQ	277
Db	151	VRNLLIVVHTIVESGVYND-----	YIQMDNSYLSNAVTFSSGESGTEGWQ	201
Qy	278	A-----VPCAPGTGPHKGVPDPVNHVLYIA--TSNWTG---PYDG---SSGDVWKFVS---	323	
Db	202	ARGSGVTVKPDVSVAYNGKYS-----LYVSGRTSNHGAQIPVDITLBEQGVKYKISVWVY	256	
Qy	324	-TSGTWTIRISVPSTDATAND-YFGVSGLTIDROHPNTIMWATQISWMPDITIIFRSTDGGA	381	
Db	257	QNSGSTQKMSLTMRORFATDPTSTYENLIYNDVPSNTWVPEPSGY-----	SIPAGV	308
Qy	382	TWTRIMDWTSPNRSRLY-----VLDTIS--AEP-----	407	
Db	309	TVSELLLYVEAQNANLAFWVDLKIYDLSKLAEPWEIPLSEIKYDYKFGVGVALSYSKI	368	
Qy	408		WLTFGVQPNPPVPSPKLGW	426
Db	369	ASDEKKMVLKHFNSITAGNEMKSELLISNNYNFSKADEFVNATSNNTAIRGHTLLW	428	
Qy	427	MDEA-----	MAIDP-----	435
Db	429	HEQTPDWFFKDANGTLLSKDALLSRLKQYIVTVGGRYKGKVVAMDVVNEAIDESQNGFR	488	
Qy	436	---FNSDRMLYGTGAILYA-----	TNDLTWKDGGQIHIAPVMVGLGETAVNDL	481
Db	489	RSNWNITCGPEYIEKAFITWAHEADPAKLFDYNDYNTENSQRQFTYNNIKLSKE-----	542	
Qy	482	ISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTT--GTSVDYAEALNPSIIVRAGS	539	
Db	543	---KGVP- IHGIG- LOCHINLWPSISEINTIKLFSSIFGLEIHIHTELDMSSFQWGSS	596	
Qy	540	FDPSQSPND-----	RHVAF-----STDGKN-----WFQ	563
Db	597	TSYSTPPRDLLIKQAMRYKELFDLFKKYNVTNTVFWGLKDDYSWLSQLNFGKSDYPLLPD	656	
Qy	564	GS-----EPGGVTTGGTV-----	AAAA	580
Db	657	GNYSKSAFVNSLIEPTVVVPVNSITLAPPAPAIQVPTSTPTPTPTVTSATPTTAPTASPA	716	
Qy	581	DGSFVWAPDGPQVYVAVFGNSWASQGVAPANAQIRS-----	DRYNP	625
Db	717	GGS--YWTPSB-----SYGALKVWYANGNSSTTNVLPNKIKITENVGTTAVDLSRVKV	767	

Qy	626	KTFTYALNSGTFYRSTDGGVTFFQP-----VRAAGLPSSGA-----VGVMFHAVP	667
Db	768	RYWYTI-DGEAAGQSVSASSINPAYIDVRVVKLGANAGGADYYVEVGFKSGAGVLAAGQS	826
Qy	668	GKEGDLWLAASSGLYHSTNGGSSWSA-----ITGVSSAYNVGFGKSAFGSSYPAVFV	719
Db	827	TKEIRLSIQKSSGSYNQSNDSYVRSANSYIENEKVTVGIDDLVWGRBPGRNAQIKWVYA	886
Qy	720	VGTIGGVTA-----YRSDDCGTT-----WVLINDCHQYGNMGQAITGDHANLR	764
Db	887	NGNLSSTNVNLNPKIKIENGVTTAVDLSRVKVRYWYTTDGEATQSVSVASSINPAYIDVR	946
Qy	765	RVYIGTNGRIVY-----GDIGGA-----PSGS-----	787
Db	947	VVKLGANAGGADYYVEVGFKSGAGVLAAGQSTKEIRLSIQKSSGSYNQSNDSYVRSANSY	1006
Qy	788	-----PSPSVSPSASP	828
Db	1007	IENEKVTVGIDDLVWGRBPGSRGKFKAG--EVTAPTPTSTPTPTPTPTPTPTPTPTPTPT	1064
Qy	829	SP	883
Db	1065	TVTATPTPTPTATPT	1124
Qy	884	PGDNOIKPGVLQVNTGSSVDLSTVTVRWFTRDGGSSTLVVNCWAAIGCGGNIRASFGS	943
Db	1125	ASTGSIKPFKIVNGSSVDLSRVKIRIYTVYVDGDKQGSV-CDWQAIGASNTVTFNVK	1183
Qy	944	VNPATPTADTYLQ	956
Db	1184	LTSGVGADYYLE	1196
RESULT 6			
T31337			
1,4-beta-glucanase (EC 3.2.1.1*) - Anaerocellum thermophilum (fragment)			
C:Species: Anaerocellum thermophilum			
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000			
C:Accession: T31337			
R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.			
Microbiology 144, 457-465, 1998			
A:Title: Properties and gene structure of a bifunctional cellulolytic enzym			
omains			
A:Reference number: Z21003; MUID:98154434			
A:Accession: T31337			
A:Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-1711 <ZVE>			
A:Cross-references: EMBL:286105; NID:e1071329; PID:e350354; PIDN:CAB06786.1			
C:Genetics:			
A:Gene: cea			
C:Keywords: glycosidase; hydrolase			

	Query Match	7.8s;	Score 398.5;	DB 2;	Length 1711;			
	Best Local Similarity	22.9s;	Pred. No. 1.7e-10;					
	Matches 197;	Conservative 116;	Mismatches 267;	Indels 281;	Gaps 40;			
Qy	262 PNNPVFWSRDGGATWQA	---	VPGAPTGFIPHKGVFDPNVNHYLYIATISNTGGPYDGS	---S	315			
Db	25 PN	----	WVRN	----	NWRGDSALKDQDNGLDLTGGWFADAGDHVF	----	NLPMSYTGTMLS	73
Qy	316 GDVWKFS	----	VTSGTWTIRISVPSTDTANDFYISGLLIDROHPNTIMVATQI	-----	365			
Db	74 WAVEYKDAFVKSGOLEHI	--	LNQIEWNDYF	-----	VKCHPSKYVYYQVGGDGSKD	124		
Qy	366 SWW	-----	PDTIIFRSTDDGATWTRINDWTSYPNRSRLRYVLDISAEPMLTFGVQP	415				
Db	125 AWWGPAEYQMQRPSFKVTSQSPGSTVYTE	----	TAASLAAASIVLK	----	DR	169		
Qy	416 NPPVPSKPLGWMDEAMAIDPENS	DRMLYGTGATLYATINDLT	KWDSGGQIHIAPMVKGLEE	475				

Db 170 NPTKAAATYLOHAK-----LYEFAEVTKSDAG-----Y 197

Qy 476 TAVNDLISPPSG-----APLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDY 526

Db 198 TAANGYNSGSGFYDELSSAAVWLYLATNDSTYLTKAE-----SYVQWPKISGNTIDY 252

Qy 527 -----AELNPSIIIRAGSFPDSSQPNDRHVAFTDGTG-----GKNWFQ 563

Db 253 KWAHCWDVINGAAL-----LLAKITGKDIYKQIIESHLDYWTGNGERIKYTPKGLAWLD 309

Qy 564 --GSEPGGVTTGTAAASAD-----GSRFVWAPGDPQPVYAVGFG-- 603

Db 310 QWGLSLRYATTAFALFYSDWVCWGPSTKKEIYKFGESQIDYALGSAGRS--FVWGFTN 367

Qy 604 -----NSWAASQGVPA-----NAQIRSD-- 621

Db 368 PPKRPHRTAHSSWADSSQSPSYHRHTLYGALVGGPGSDSDSYTDDISNYVNEVACDYNA 427

Qy 622 -----RVNP-----KTFYALNSGTFFYRSTDGGVTFFQPVAAAGLPSGGAAGYWFH 664

Db 428 GFVGALAKMYQLYGGNPIPDFKAIETPTNDEFF-----VEAGINASGTNFIK 476

Qy 665 AVPGKEGDLMAASS-----GLYHSTNGGSSAITGVSAVNVGFKSAP----- 710

Db 477 AIVNQSG-WPAKATDKLFYFVDLSLSELIKAGYSPNQLT-LSTNYNQAGKVSQPYVWDA 534

Qy 711 --GSSYPAVFVGTI--GGVTGAYRSDDCGTTWVLIINDHQHGVNMGQAITGDHANLRV 766

Db 535 SKNTYYILVDFTGLIYPGGQDYKKE-----VQFRIAAPQNVQWDSNDYSFQDIKGV 588

Qy 767 YIGT-----NGRGIVYDGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 816

Db 589 SSGSVWTKIPIYLDGVKVGDPGFGTSGATPTPTA-----TATPTPTPTPTPTPTPT 642

Qy 817 SSSPSSP 875

Db 643 TS--TATPTPTPTPTPTPTPTPTPTPTA--TPTSTPTPTSP-----TPSSTPTVAGGQIKV 690

Qy 876 QYKNNDSAPGDNQKPGLOVVNTGSSSVLDSTVTVYRWFTRDGGSSLYVNCDAATGCG 935

Db 691 LYANKETNTNTIRPWLKVNTGSSSIDLSRVTVYRWFTRDGGSSLYVNCDAATGCG 935

Qy 936 NIRASFGSVNATPTADTYLQ 956

Db 750 NVTEKFVKLSSSVSGADYYLE 770

RESULT 7

T17120

cellulase (EC 3.2.1.1) precursor, thermoactive - Caldocellum saccharolyticum

C:Species: Caldocellum saccharolyticum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000

C:Accession: T17120; A43745

R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.

A:Title: Microbiol. Biotechnol. 43, 291-296, 1995

A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophilic bacterium Caldocellum saccharolyticum

A:Reference number: 218698; MUID:95336703

A:Accession: T17120

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1742 <TEO>

A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1

R:Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 57, 694-700, 1991

A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a multidomain cellulase from the extreme thermophilic bacterium Caldocellum saccharolyticum

A:Reference number: A43745; MUID:91247819

A:Accession: A43745

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1516-1544, 'A', 1546-1742 <LUE>

A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293

C:Genetics:

A:Gene: cels

C;Keywords: glycosidase; hydrolase

Query Match 7.7%; Score 397.5; DB 2; Length 1742;

Best Local Similarity 21.7%; Pred. No. 1.9e-10;

Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

Qy 262 PNPVFRSGGGATWQA-----VPGAPGTGPHKGVDPVNVHLY-IATSNVTGGPYDSSG 316

Db 48 PN-----WVRN-----NWRGDSALKDQDGLDTGCGWFDAGDHVAFNLPMSTGTGTMLSWAA 100

Qy 317 DVKSFVSTGVTWTRISPVSTDTANDYFGYSLGTIDRQHPNTIMVATQI-----SMW- 368

Db 101 EYKDAFVKSQLEHI--LNQIEWNDIF-----VKCHSKYIYVQVGGGKDHAWMG 151

Qy 369 -----PDTIIFRSTGGATWTRIMWTSTYPNRSRLRYVLDISARPEWLTFTGQPNPPV 419

Db 152 PAEVMQMERPSFKVTQSSPGSA-----VVAETAASLAASIVLKDRN 193

Qy 420 PSPKLHMDAMALDPNSDRMLYGTGATLYATNNDLTKWDSGGQIHIAPMVKGLEETAVN 479

Db 194 PTAAATYLOHA-----KDIYFAEVTKSDG-----YTAAN 224

Qy 480 DLISPPSG-----APLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDY-- 526

Db 225 GYNSWSGFFYDELSSAAVWLYLATNDSTYLTKAE-----SYVQWPKISGNTIDYKWAH 279

Qy 527 -----AELNPSIIIRAGSFPDSSQPNDRHVAFTDGTG-----GKNW----- 561

Db 280 CWDDVHNGAAL--LLAKITDKDYKQIIESHLDYWTGNGERIKYTPKGLAWLDQWS 336

Qy 562 -----FQSEPGGVTTGTTVAASADG-SRFVWAPGDPQPVYAVGFG----- 603

Db 337 LRYATTTAFALFYSDWVCWGPSTKKEIYKFGESQIDYALGSAGRS--FVWGFTNPKR 394

Qy 604 -----NSWAASQGVPA-----NAQIRSD----- 621

Db 395 PHRTAHSSWADSSQSPSYHRHTLYGALVGGPGSDSDSYTDDISNYVNEVACDYNAGFVG 454

Qy 622 -----RVNP-----KTFYALNSGTFFYRSTDGGVTFFQPVAAAGLPSGGAAGYWFH 668

Db 455 ALAKMYLLYGGNPIPDFKAIETPTNDEFF-----VEAGINASGTNFIKAI-- 501

Qy 669 KEGDLWLAASSGLYHSTNGGSSWA-----ITGVSSAVNVGFKSAPGSSYPAVF 718

Db 502 -----VNNOSGWPATNKLKFRYFVDLSLSELIKAGYS-----PNQL 537

Qy 719 VVGTI-----GGVTGAYRSDDCGTTWVLI-----DQHOYGNMGQAITGDHANLR 765

Db 538 TLSTNYNQAGKVSQPYVWDSRNIIYIYLDFTGLIYPGGQDYKKEVQFRIAPQNVQ- 596

Qy 766 VYGTNGRGIVYDGGAPSGS-----PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 807

Db 597 --WDNSNDYSFQDIKGVSSGVVTKYIPLYEDIKVWGEPEGTS-----GVSPTPTAS 648

Qy 808 SSSPSS 867

Db 649 VTPTPTPTPTA-----TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 700

Qy 868 P-VSGGVKVOYKKNDSAPGDNQKPGLOVVNTGSSSVLDSTVTVYRWFTRDGGSSLYV 926

Db 701 PATSGKIKVLYANKETNTNTIRPWLKVNTGSSSIDLSRVTVYRWFTRDGGSSLYV 926

Qy 927 CDWAAICGNIRASFGSVNATPTADTYLQ 956

Db 761 -DWAQIGASNVTFKFKVLSVSSVSGADYYLE 789

RESULT 8

S50755

hypothetical protein VSP-3 - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii

C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000

A::Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A::Reference number: A48954; MUID: 93119139
A::Accession: A48954
A::Status: preliminary
A::Molecule type: nucleic acid
A::Residues: 1-1331 <GB>
A::Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
A::Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)
R::Luetli, E.; Bhana Jasmnat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A::Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding
A::Reference number: A43745; MUID: 91247819
A::Accession: B43745
A::Status: preliminary
A::Molecule type: DNA
A::Residues: 1-337, 'P'PQHQHQ' <LUE>
A::Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A::Note: the authors translated the codon CAC for residue 262 as Glu
A::Note: this sequence has been revised in reference A48954
C::Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 6.6%; Score 340.5; DB 2; Length 1331;
Best Local Similarity 22.7%; Pred. No. 5e-08;
Matches 140; Conservative 97; Mismatches 189; Indels 191; Gaps 29;

Qy	426	WMDPAMAI	DPSNDRMLYGTGATLYATNDLTKWDSSGOATHIAPMVKLBETA VNDLISPP	488
Db	141	WKETKSVDLG-	NEDFVILINIGNEPYGNNNQN-----VNDTKN--	178
Qy	486	SGAPLISALGDGLGFTHADVTAPS-	TIFTSPVETTGSVDYAELNPSSIIVRA---GSF	540
Db	179	----AIALRD-AGFKHTIMVDAPNWQDWSNTMRNAOSIMEADPLRNLIFSIIHWGYV	233	
Qy	541	DPSSQPNDRHVFSTDG----	GKNWFQGSEPGGVTTGGTVAASADGRFVWPACPDGPQ	595
Db	234	NTASKVEEYIKSFVKDLPLVIFEGFHQTD-	-----GDPDEE	270
Qy	596	VV-----YAVG-FGNSWAASOGVPANAOLRS-	DRVNPKTFFALSNGTFYRSTDGG---	643
Db	271	AIVRYAKOYKI GLFSWSCGNSSYVG YLDVMNMDDPNPTPW-----GWMYKTNAIGTSS	325	
Qy	644	-----VTFQPVAAALP-----SSGAVGMFPHPVPGKEGLD-	673	
Db	326	TPPTPTSVTPTPTPTPTPTTVATPTPTPTPVSTPATSQIKVLV---	-ANKETNSTTNT	382
Qy	674	--WL-----RASSCL-----YISTNGCGSSWALTGVS---	SAYNVGF-----GKSAP	710
Db	383	IRPLKLVNYSGCCSIDLSRVTRIYWTVTDGERAQSAISDMAQAICASNVTKFVKLSVVSS	442	
Qy	711	GSSYPAVFVVGTIGGVGTAYSRDCCGTWTWLI NDD--OHQYCN--	WGQAITGDHANLR	764
Db	443	GADY--YLETGFGSKAGCOLQP GKDTGEIQMRFNKDDWSNYNQDNWSWIQSMTS-	-----494	
Qy	765	RVIYICTNRGI VYG D----ICGAPGPSPPSPSVSPASPSLPSPSSPSPPSPSPSSPP	820	
Db	495	---YGENEKVTAYIDGVLVWQZEPSCGA-----TPAPAP	524	
Qy	821	SSSP	879	
Db	525	TATPT	576	
Qy	880	NDSAPGONQIKPKQLQVYNTGSSSDLSSTVTRYFYMFTRDGGSSLVYNCDMAAII CGGNIRA	939	
Db	577	KETNSTNTIRP LKLVNYSGCCSIDLSRVTRI WYTVTDGERAQSAIS-DWAQI GASNVT F	635	
Qy	940	SFGSVNPATPTADTYLQ	956	
Db	636	KFVKLSSSVSGADYYLE	652	

RESULT 11
S76211

hypothetical protein slr0442 - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76211
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasunaga, K. *Proc Natl Acad Sci USA* 95:1367-1372, 1998
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A:Reference number: S74322; MUID:97061201
A:Accession: S76211
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-611 <RAN>
A:Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BAA18470.1; PID:d101
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 6.0%; Score 308.5; DB 2; Length 611;
Best Local Similarity 26.7%; Pred. NO. 5.6e-07;
Matches 142; Conservative 51; Mismatches 180; Indels 159

[illegible]

RESULT 12

S/5138
hypothetical protein slr1753 - *Synechocystis* sp. (strain PCC 6803)

C:Species: Synechocystis sp.
C/Variety: PCC 6803
A:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C/Accession: S75138
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, N.; Kanehisa, A.; Kohno, K.; Takaiwa, F.; Terashima, T.; Uemura, M.; Wada, K.; Watanabe, Y.; Yoshida, K.; Zutschi, J.

A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.s.
A:Reference number: S74322; MUID:97061201

A:Accession: S75138
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1749 <KAN>
 A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BAAL7052.1; PID:d101778
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG

```

Query Match          5.9%; Score 301.5; DB 2; Length 1749;
Best Local Similarity 19.0%; Pred. No. 3.9e-06;
Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;

QY 17 VSLAATASFAAAALGVLPIAITASPAHAATQPYTWSNVAIGG-----62
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 23 LALLACSSFSFG----NVLAQNTIPADPGTCTTVDAGNQFNIGGSLGQGNLFSLQ 78
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 63 -----GFVDGIVFNEGAPGILYVRTDIGMYRWDA 92
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 79 QFGLDQGIANFLSNPDIRNLTIRIVGGDASIIINGLIQVSGGNANFLMNPAGMIFGPN 138
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 93 A-----NCRWPLLDVGNWNGVNVVS-----116
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 139 SINVPGEFVVTGSAIGFNDQMFQV-----FSDNDYNALIGNPSQAFDLANPGLIINA 193
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 117 -----IAADPINTNKWAAVGMVYNSWDNDGAILRSS-----149
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 194 GDLVTEGKNLTFLAGNIVNTGSLAAPGNTVAAVPQGNRIRISOAGSLLSLEVEVSPQ 253
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 150 -DQATWQITPLPKLGGNPGRMGRERLAVDPNND-----NLLYFGASGKGLWRSTD 202
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 254 MNQGSFSLDLPLTLLQGASNLDLG--LAVQPNVSGVTNGTNALVPLPGSVTISQNV 311
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 203 SGATWSQMTNFPDVGTVIANPTDTTGYS--DIO-----GVVWVAFD-KSSSSLGOASK 253
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 312 AS---GKSTNISS--GGQVATAGDIAVQATVDVSGNGGVTFRIGDFGQILPLNASQ 367
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 254 TIFGVADPNPV-----FWSRDGGA--TWQAVPGAPTGFIPHKGVFDPVNVHLYIATSN 307
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 368 TLI-----DSNVKADALLGNGGTIVWADDSFRSGNI-----SAQ 406
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 308 GGPYDSSGDVWKFVSVSGTWTRISVPVSTDTANDYFGYSLGILIDROHPNTIMVATQISW 367
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 407 GGTMGNGG---FVETSG-----AKSLMWDDTARVNTFATMGELGT 444
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 368 W---PDRITIFRSTGGATWTRIMD---WTSYPNRSRYVLDISAEPLWTFGVQNPVPV 420
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 445 WLLDPLEIIVGTTDLLADPKLVSVLTITSLDNGVILQADQSIQVQANFSADPSAP--502
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 421 SPKLGWDEAMADPPNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHI--466
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 503 -GNLTFDSPITIDALFS---LGTSGIIFANTGPINTGNTLVTSPTNLDLFONKIQIQLNA 557
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 467 -----APWKGLETAVN---DLISPPS-----GAPLISALGDLG-----498
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 558 NTFITAPGYDIYPRKSVNGGFDLLGNANFYFDGAGITTPKLSFGVTATEIYVGNDIVT 617
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 499 -----GFTHADVT-AVSTIFTSPVFTTG-----TSVDYAEPLNP 531
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 618 QGNQIFDGVYGLQPVNLTSAGSVITFNILLNGSLQVTOANTIVSQPSSSLSAVEIAS 677
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 532 SIIVRAGSFDPSPQNDHRVAF---STDGKNMFQSEPGGVTTG-----GTVA 577
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 678 DVLLNAG-----QNVSFGNINTRGGNVDIQL--GNISTGSIIVTSFFGNAGNVI 725
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 578 ASADG-----SRVWAPGDPQVYVAVFGNSWA-----ASQGV 612
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 726 LNAGGTLTTGYIETSGTNGGDDVTSSGNTSTAYIDTRGFGDGLIDSLGGAVSIESKGD 785
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 613 PANQAIFSRDVRNPKTFVALSNGTIFRSTDGGVTFPQVAAGLPSSGAVGVMMFHAVPGKE--670
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 786 ITTAFIDTGAYSIESFNEGTGGNVLFTADGSITNTYIFTAGKNGG--DIFFQAGESIEII 843
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  
```

```

QY 671 -----GDLWLAA-----SSGLYHSTNGGS-----SWSAITGV 697
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 844 DYLNVTYGSQTSQSDVYVEAPLDLISGSIYIYGGGEPGNVFLQAGGDITTSYIDTSAANGG 903
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 698 SSAVNVG-----FGKSAPGSSYPAFVYVGT 722
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 904 DIFQSGGDEVGVFTFKGYEGRGDGVYVETRYFRAIDGFLGCEGPFVYTAGLTVG-962
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 723 IGVVTGAYRSDDCGTTWVLINDQHOYGNNGQAITGDHANLRVYIGTNGR--GIVYGDI 780
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 963 -----GSVYI-----FGGSEPFIIIGNPI-----TNGTIGAISSGDD 994
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 781 GGAPSGSP-----SPSVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 825
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 995 NTPVICTGTPPIEDFTLDNIITITEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1054
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 826 PSRSPSPSPSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1055 PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1097
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  
```

RESULT 13

TI0361
 hypothetical protein 92 - Orgyia pseudotsugata nuclear polyhedrosis virus
 C:Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNVP
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T10361
 R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.
 Virology 229, 381-399, 1997
 A:Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedro
 A:Reference number: Z17011; MUID:97271300
 A:Accession: T10361
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-279 <HR>
 A:Cross-references: EMBL:U75930; NID:g2934903; PIDN:AACS9091.1; PID:g1911338
 C:Superfamily: proline-rich protein 3

Query Match 5.7%; Score 293; DB 2; Length 279;
 Best Local Similarity 55.8%; Pred. No. 1.1e-06;
 Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;

```

QY 783 APGSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 842
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 78 SPRTPTSLSPRTPTSPRTPTSPRTPTSPRTPTSPRTPTSPRTPTSPRTPTSPRTPTSP 137
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 843 SPSPSSSPSSSPSSSPSSSPSPSSSP 868
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 138 SPTPTPTPTPTPTPTPTPTPTPTPT 163
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  
```

RESULT 14

S50754
 hypothetical protein WP6 - Chlamydomonas eugametos
 C:Species: Chlamydomonas eugametos
 C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
 C:Accession: S50754
 R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Kils, F.M.; Goodenough, U.W.; Hari
 Plant Mol. Biol. 26, 947-960, 1994
 A:Title: Domain conservation in several volvocalean cell wall proteins.
 A:Reference number: S50754; MUID:95093034
 A:Accession: S50754
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-351 <WOE>
 A:Cross-references: EMBL:L29028; NID:g530877; PIDN:AAB53954.1; PID:g530878

Query Match 5.7%; Score 291; DB 2; Length 351;
 Best Local Similarity 64.4%; Pred. No. 1.8e-06;
 Matches 58; Conservative 10; Mismatches 18; Indels 4; Gaps 2;

QY 783 APSCSPSPVSPASPSLSPSPSPSPSPSP--SPSSSPSPSPSPSPSPSPSPSPSPSP 840
Db 175 SP 234
QY 841 SASP--SPSSSP 868
Db 235 VASQSP 264

RESULT 15
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
A:Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GE>
A:Cross-references: EMBL:U08046; PIDN:AC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: x
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 5.7%; Score 290.5; DB 2; Length 2232;
Best Local Similarity 21.8%; Pred. No. 1.6e-05;
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;

QY 18 SLLAATASFAVAALGVLPTA---ITASPAHAATTQPYTHSNVAI--GGGFFVDGIVFNE 72
Db 235 SLGTSSSLPSSISTALPIASSASSSSPAASSTPVLSSSTIQSSSGSTFPSSVASSP 294

QY 73 GAPGILYVRTDIGMYRWDAANGRWIPLLDWGVNNGWYGVVSIAD-----PINTNK 126
Db 295 STVG-----STSG-----AASSSYATVSTIAGSTGSTITPVPSS 330

QY 127 VWAAVGYMTNSWDPNDGAILR--SSOGATWQITP-LPFKLGGNMP---GRGMGERLAVD 180
Db 331 --STIGSSTPSASSSSSGTMTISGSGTGTVPVPGSSSTFASSTPIASSSPGSGTVA 388

QY 181 PNNDNILYFGAPSGKGL---WRSTDGATWSOMTNPDPVCTYIANPTDTGYOSDIOGVV 237
Db 389 PGSSSTYGSSTPSASSSSSGTMTNSGSGTGTVPVAPVSSSTFGSSTPIASSSSSGSTVT 448

QY 238 WVAFDK-----SSSLGQAS-----KTFVGVADPNPNPVFVWSRDGGATWQA 278
Db 449 VVSGSSSTYGSSTPSASSSSAGTASTISGSGTGTATVPSS-----SSVGSSTQSA 500

QY 279 VPGAPTFIPHKGVDFDVNHLVIAFNTGCGPYDGGSDGVKFSVTGWTIRISPVPSD 338
Db 501 SPSSPG-----TMTSVSGP-TGSTVTVVPGSSTSPA-PSSSPNPPSS 540

QY 339 TANDYFGYGLTIDRQHPNTIWATOISWMPDTHFRSTDGATWT---RIWDVTSYPNR 395
Db 541 PAS-----TGSTITIGSSSIIIVST-----VSGSTVSGSGTGTQSQTSLASSTATPGS 586

QY 396 SLRYVLDISAE-----PWLTFGVQPNP---PVPSFKLGWMDMAIDPFNS 438
Db 587 SSVVPSSSPQSPSPAPNTGTTPTSQTSQSPSPSMNPSSSTPTGSSQSTITPEGSTA 646

QY 439 DRMLYGTGATLYATNDLTKWDSGGQTHIAPWVKLEETAVNDLISPPSGAPLISALGDL- 497
Db 647 SSPTGTGSGTFSVATEVTSQST-----VPSGSSLGTTQSTNNSPSPSLSPSTSGMSTLT 700

QY 498 -----GGTFHADVTAVPS-----TIFTSPVFTTGTSDVYAEINPILVIRAGS 539

Search completed: July 2, 2002, 09:13:57
Job time: 125 sec

Db 701 SEPSPSTOSSGAQSTLTTPSPNPSPSTSSLESSTSGAITSAGTTMTSPSQSSVGS 760
QY 540 FDPSSQPNDRHVAFTDGGKNWFGQS--EPG-GVTTGGTVAASADGSRFVWAPG----- 590
Db 761 SQGSTSP-----AASITSGEMTSQGSTQTPGSSVTSAAILTSTQOSVSTNSPGSTVTRP 815
QY 591 -----DPCQPVVY-----AVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGTFFYS 639
Db 816 STVSGSTSSGSTVTVGSTEASTSGSSVASSPAPSTSQ-----NPNPSTSSGSSMITQS 869
QY 640 TDGGVTQPVAAGLPSSGAVGVMFHAPGKGGDLWLAASSGLYHSTNGGSSWSA-----IT 695
Db 870 PYPQSSTSPVESSTTFS-----PGSPGTLTSTSPSPSQSTTIGSTQGSTSPGIS 919
QY 696 GVSAYNVGVFGKSPAGSSSYPAFVVGVTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQA 755
Db 920 TTSEEMTSQGSTQTPGSGTGTVPSTVSDST----- 951
QY 756 ITGDHANLRVYIGTNGRGIVYGDIGAGPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSP 810
Db 952 -----SSGSTVTVGSTE-----SSSIPSTSONTPNPSSTSGSSMSTQTQSSQ 995
QY 811 SPSP--SPSSSPSSSPSP-----SPSPSPSP-----SRSP-----SPSASPSPS 848
Db 996 STSPVESSTSGATSSGSGPCTTLTSPSPSPSTTIGSSQSGSTSPVVSTISQGSTETPGS 1055
QY 849 SPSPSSSP-----SSSPSTPSSSPVSGGVKVVQYKNN 881
Db 1056 TGTVTNKPSVSGSSGSTATMGSTEASSTSGSSTSPNPSQSTSPSTSGA-----T 1108

QY 882 SAPGDNQIKPGLQVNTGSSVDLSTVTYVWYTRDGGSTLYNC--DMAAICGNGIRA 939
Db 1109 SSPGSS-----GTLTSTSPSPSQSSTIG-----SSOGSTSPVVSTTSGDMTSQGSTQIPG 1159

QY 940 SFGS--VNPATPTADT 953
Db 1160 STGSTVTPSTGSGST 1175

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 28.05 Seconds
(without alignments)
1321.020 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	397.5	7.7	1742	1 GUNA_CALSA	P22534 caldocellum
2	340.5	6.6	1331	1 MAND_CALSA	P22533 caldocellum
3	293	5.7	279	1 Y01 NPVOP	O10341 orgyia pseu
4	290	5.6	1039	1 GUNB_CALSA	P10474 c endodluca
5	283.5	5.5	474	1 VTP3_TTV1V	P19275 thermoprote
6	272	5.3	555	1 GPL_CHURE	Q9fpg6 chlamydomon
7	261.5	5.1	1853	1 CIPA_CLOTM	Q06851 clostridium
8	250	4.9	268	1 NO20_MEDTR	P93329 medicago tr
9	245	4.8	772	1 CIPB_CLOTM	Q01866 clostridium
10	245	4.8	879	1 GUN1_CLOTM	Q02934 clostridium
11	244	4.8	1664	1 SLP1_CLOTM	Q06852 clostridium
12	235.5	4.6	2124	1 PGCA_RAT	P07897 rattus norv
13	222.5	4.3	360	1 VTPX_TTV1	P19274 thermoprote
14	221	4.3	2132	1 PGCA_MOUSE	Q61282 mus musculu
15	218.5	4.3	3164	1 TEGU_HSV11	P10220 herpes simp
16	216	4.2	449	1 APG_BRANA	P40603 brassica na
17	214	4.2	532	1 SPG7_DICDI	P22698 dictyosteli
18	210.5	4.1	562	1 GUN1_ACICE	P54583 acidothermu
19	209	4.1	535	1 SPKC_SYNY3	P74745 synecocyst
20	209	4.1	2415	1 PGCA_HUMAN	P16112 homo sapien
21	208.5	4.1	485	1 SSGP_VOLCA	P21997 volvox cart
22	207.5	4.0	544	1 G1Q0_DICDI	Q06885 dictyosteli
23	205.5	4.0	534	1 APG_ARATH	P40602 arabidopsis
24	205.5	4.0	700	1 GUNA_PAELA	P29719 paenibacill
25	204.5	4.0	499	1 GUN2_BACSU	P10475 bacillus su
26	204.5	4.0	1970	1 RPBI_HUMAN	P24928 homo sapien
27	204.5	4.0	1970	1 RPBI_MOUSE	P08775 mus musculu
28	204	4.0	444	1 GUNN_ERWCA	Q59394 erwinia car
29	203.5	4.0	499	1 GUN1_BACSU	P07983 bacillus su
30	201	3.9	456	1 GUNA_MICBI	P26414 microbispor
31	199.5	3.9	499	1 GUN3_BACSU	P23549 bacillus su
32	199.5	3.9	1367	1 AMVH_YEAST	P08640 saccharomyc
33	199	3.9	1324	1 IRS2_HUMAN	Q9y4h2 homo sapien

ALIGNMENTS

RESULT 1

ID	GUNA_CALSA	STANDARD;	PRT;	1742 AA.
AC	P22534;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)			
DE	(Cellulase A).			
GN	CELA.			
OS	Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Thermoanaerobacter group; Caldicellulosiruptor.			
OX	NCBI_TaxID=44001;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=95336703; PubMed=7612247;			
RA	Te'O V.S., Saul D.J., Bergquist P.L.;			
RT	"cela", another gene coding for a multidomain cellulase from the			
RT	extreme thermophile Caldocellum saccharolyticum.;			
RL	Appl. Microbiol. Biotechnol. 43:291-296(1995).			
RN	[2]			
RP	SEQUENCE OF 1516-1742 FROM N.A.			
RP	MEDLINE=91247819; PubMed=2039230;			
RA	Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;			
RT	"Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for a beta-mannanase from the extremely thermophilic			
RT	bacterium 'Caldocellum saccharolyticum'.			
RL	Appl. Environ. Microbiol. 57:694-700(1991).			
CC	-I- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN			
CC	ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL			
CC	DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE			
CC	CELLULOSE.			
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-I- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL			
CC	SITE FOR PROTEOLYSIS.			
CC	-I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	E (FAMILY 9 OF GLYCOSYL HYDROLASES).			
CC	-I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	L (FAMILY 48 OF GLYCOSYL HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L32742; AAA91086.1; -			
DR	EMBL; M36063; AAA72860.1; -			
DR	EMBL; L01257; -; NOT_ANNOTATED_CDS.			
DR	PIR; A43745; A43745.			
DR	HSSP; P26221; 1TF4.			
DR	InterPro; IPR001956; CBD_3.			

Q47096 erwinia car
P11414 cricetus
P16356 caenorhabdi
Q28343 canis famil
O13368 candida alb
P12021 sus scrofa
P46590 candida alb
P35084 dictyosteli
P13088 bos taurus
P02893 plasmodium
P18616 arabidopsis
P19597 plasmodium


```
DR InterPro: IPR000556; Glyco_hydro_48.
DR InterPro: IPR001701; Glyco_hydro_9.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR00844; GLHDLRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 7.7%; Score 397.5; DB 1; Length 1742;
Best Local Similarity 21.7%; Pred. No. 4.9e-10;
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

QY 262 PNNPVFWSRGGATWQA---VPGAPTGFPHKGVDPPVNHVLY-IATSNVTGGPYDGGSSG 316
DB 48 PN-----VWRN---NWRGDSALKDQDNGLDLTGWFEDAGDHVKFNLPMSYTGTHLSWAAY 100
QY 317 DVNKFVSTGTRISPVSTFDTANDFYGYSGLTIDRQHPNTIMVATQI-----SWW- 368
DB 101 EYKDAFVKSGLEHI--LNOIEWNDYF-----VKCHPSKVVYVYQVGGDGKDHAWWG 151
QY 369 -----PDTIFIRSTDGGATWTRWDWTSYPNSRLRVLDISAEPWLTFGVQNPVPV 419
DB 152 PAEVMQMERPSFKVTQSSFGSA-----VVAETAASLAASASIVLKDKN 193
QY 420 PPSKLGMDMAIDAIDFNSDRMLYGTGATLYATNDLTKWDSGQIHIAPMKVGLLEETAVN 479
DB 194 PTRAAATYLOHA-----KDIYEFAEVTKSDSG-----YTAAN 224
QY 480 DLISPPSG-----APLISALGLDGFETHADVTAVPSTFTSPVFTTGTSDVY---- 526
DB 225 GYVNSWGFYDELSWAAVWLYLATNDSYLTAKAE-----SYVQNPWKISGNSLIDYKWAH 279
QY 527 -----AELNPSIIIRAGSEDPSSQPNDRHVAFSTDG-----GKNW----- 561
DB 280 CWDDVHNGAAL---LKAKTDRKYQIIESHLDYWTGTVNGRIKYYTKPKGLAWLDQWSG 336
QY 562 -----FOGSEPPGVYTTGGTVAASADG--SREVMAPGDPQPVYAVGFG----- 603
DB 337 LRVATTATFLAFYSDWSGCPGCKEYRKFGESQIDYALGSTGRS--FVWFGTNPVKR 394
QY 604 -----NSWAASQGVPA-----NAQIRSD----- 621
DB 395 PHRTAHSSWADSQSIPSVYHRHTLYGALVGGPGSDSDSYTDDISYVNVNEVACDYNAGFVG 454
QY 622 -----RVNP-----KTFYALNSGTFYRSTDGGYTFQPVAGLPSSGAVGMFHAVPG 668
DB 455 ALAKMYLLVGNPNIPDKAETETNDEFF-----VEAGINASGTFNTEIKAI-- 501
QY 669 KEGDLMAASSGLYHSTNGSSWSA-----ITGVSSAVNVGVFGKSAPGSSYPAVEF 718
DB 502 -----VNNQSGWPATNKLKFRYFVDLSSELKAGYS,-----PNQL 537
QY 719 VVGTYI-----GGVTGAYRSDCGTGTWVLIND-----DQHQYGNWGOAITGDHANLRR 765
DB 537 -----DQHQYGNWGOAITGDHANLRR 765
```



```
DR EMBL: M36063; AAA72861.1; --
DR PIR: B43745; B43745.
DR PIR: A48954; A48954.
DR HSP: Q06851; INBC.
DR InterPro: IPR001956; CBD.3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD.3; 2.
DR Pfam: PF00150; cellulase; 1.
DR PRODOM: PD001947; CBD.3; 2.
DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Hydrolase; Glycosidase; Cellulose degradation; Signal;
KW Multifunctional enzyme.
FT SIGNAL 1 41
FT CHAIN 42 1331
FT DOMAIN 42 325
FT DOMAIN 326 361
FT DOMAIN 362 518
FT DOMAIN 519 564
FT DOMAIN 565 720
FT DOMAIN 721 780
FT DOMAIN 781 1331
FT ACT_SITE 162 162
FT ACT_SITE 257 257
FT CONFLICT 338 338
FT CONFLICT 340 346
FT CONFLICT 1331 AA; 146892 MW; FFBCA51BB8D8F0E0 CRC64;
SQ SEQUENCE 1331 AA; 146892 MW; FFBCA51BB8D8F0E0 CRC64;

Query Match 6.6%; Score 340.5; DB 1; Length 1331;
Best Local Similarity 22.7%; Pred. No. 9.6e-08;
Matches 140; Conservative 97; Mismatches 189; Indels 191; Gaps 29;

QY 426 WMDAMAIIDFNSDRMLYGCATLYATNLDLTKWDSGQIHAPMKGLEATVNDLISPP 485
DB 141 WKEIKSLDVG-NEDFVITINIGNEPYGNNTYQNW-----VNDTKN-- 178
QY 486 SGAPLISALGDGLGTHADVTAPVS--TIFSPVFTTGSVDYAEALNPSIIVRA---GSF 540
DB 179 ----AIKALRD-AGFKHTIVMDAPNWDQNSNTMRDQAQSIEMADPLRLNLFVSHMYGVY 233
QY 541 DPSSQPNDRHVAFTSDG-----GKNWFQSGEPGGVTTGGTVAASADGSRFVWAPGDPGP 595
DB 234 NTASKVEEYIKSFVDKGLPLVIGFEGHQT-----GDPDEE 270
QY 596 VV-----YAVG-FGNSWAASQGVPAQAQIRS--DRYNPKTFYALNSGTFRSTDDG--- 643
DB 271 AIVRYAKQYKIGLFSWCGNSSVYGYLDMVNNMDPNNTPTW-----GQYKTNIAIGTSS 325
QY 644 -----VTFQVAAAGLP-----SSGAVGVYMFHAPQCKEGLD--- 673
DB 326 TPTPTSTVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 382
QY 674 ---WL-----AASSGL-----YHSTNGSSWSAITGVSV--SAVNVGF-----GKSAP 710
DB 383 IRPLKLVNNGSSSIDLSRVTIRVYTVVDGERAQSAISDWAQIGASNVTFKFKVLSVSVS 442
QY 711 GSSYPAVFVVGTTGGVTVGAYRSDCGTTWTLINDD---HQYQGN---WGQAITGDHANLR 764
DB 443 GADY--YLEIGFGKSGAQOLQPGKDTGTGIQMRFNKDDWSNTNOGNDWSWISQMTS----- 494
QY 765 RVYIGTNGRGIYGD---IGGAPSGSPSPSVSPSASPSLSPPSPSSPSPPSPSPSSSP 820
DB 495 ---YGENEKTAYIDGVLVWQEPGSA-----TPAPAP 524
QY 821 SSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 879
DB 525 TATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 576
QY 880 NDSAPGDNQIKPGQLQVNTGSSSVSLTVTVRYWFTTRDGGSTLVYVNCDAWAAICGNIRA 939
DB 577 KETNSTNTIRPMLKVNNGSSSIDLSRVTIRVYTVVDGERAQSAIS--DNAQIGASNVTF 635
QY 940 SFGSVNPTATTADTYLQ 956

Db 636 KFKVLSVSSVSGADYILE 652

RESULT 3
Y091_NPVOP
ID Y091_NPVOP STANDARD; PRT; 279 AA.
AC Q10341;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 29.3 kDa protein (ORF92).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpmNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75930; AAC59091.1; --
CC SEQUENCE 279 AA; 29289 MW; 6FA4DAA01009DBF0 CRC64;

Query Match 5.7%; Score 293; DB 1; Length 279;
Best Local Similarity 55.8%; Pred. No. 1.9e-06;
Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;

QY 783 APSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 842
DB 78 SPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 137
QY 843 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
DB 138 SPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 163

RESULT 4
GUNB_CALSA
ID GUNB_CALSA STANDARD; PRT; 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
(BC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
(Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
(1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocolium saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoclostridium group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
RT "Nucleotide sequence of a gene from Caldocolium saccharolyticum"
```



```

DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
DE glycoprotein 1).
DE GN
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OC NCBI_TaxID=3055;
RN [1]
RN SEQUENCE FROM N.A.
RX PubMed=11258910;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RA "Glycosylated polyproline II rods-with-kinks as a structural motif in
RA plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RN PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RA "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RA encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -!- SUBUNIT: Associates with GP2 and GP3.
CC -!- PTM: N-glycosylated and O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF309494; AAG45420.1; -.
DR EMBL; M58496; AAA69706.1; ALT_SEQ.
DR GlycoSuiteDB; Q9FP06; -.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR Glycoprotein; Repeat; Signal.
KW SIGNAL 1 29
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
SQ
Query Match 5.3%; Score 272; DB 1; Length 555;
Best Local Similarity 44.6%; Pred. No. 3.le-05;
Matches 74; Conservative 15; Mismatches 41; Indels 36; Gaps 8;
QY 784 PSGSPSPSVSPS-ASPSLSGSPSPSPSS---SPSPSPSPSS-----SPSSSPSPS---PSPSPS 833
DB 287 PPSPSPSPSPAPPTPTPTSPSPSPVPVPSAPVPVPSAPPSPAPSPPPSPAPTSPS 346
QY 834 PSRSPSPASPSGSPSPSPSSSPSSSPSPSSPSGSGVKVQYKNDSAPGDNQIKPL 893
DB 347 PPSPPSPSPSPSPSPSPSPSP1PSPPKPSPSFVA--VKLVWADIAFD----- 397
QY 894 QVWNTGSSVDLSTVTYRFTRDGSGSTLYNCWAAIGC-GNIR 938
DB 398 --LNGTIS-----TRGSASRWGEPDIAGTKCKGNLK 427
RESULT 7

```


[illegible]

FT	DOMAIN	486	580	G2-B.
FT	DOMAIN	587	682	G2-B'.
FT	DOMAIN	685	798	KS.
FT	DOMAIN	801	1226	CS-1.
FT	DOMAIN	1227	1909	CS-2.
FT	DOMAIN	1910	2124	G3.
FT	DISULFID	51	133	BY SIMILARITY.
FT	DISULFID	175	246	BY SIMILARITY.
FT	DISULFID	199	220	BY SIMILARITY.
FT	DISULFID	273	348	BY SIMILARITY.
FT	DISULFID	297	318	BY SIMILARITY.
FT	DISULFID	509	580	BY SIMILARITY.
FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1914	1925	BY SIMILARITY.
FT	DISULFID	1942	2034	BY SIMILARITY.
FT	DISULFID	2010	2026	BY SIMILARITY.
FT	DISULFID	2041	2084	BY SIMILARITY.
FT	DISULFID	2070	2097	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1842	1842	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;		

Query Match 4.6%; Score 235.5; DB 1; Length 2124;
Best Local Similarity 21.3%; Pred. No. 0.0044;
Matches 224; Conservative 115; Mismatches 359; Indels 353; Gaps 48;

QY	112	NGVSTAAADPINTNKVAAVGMVNTSNWDPNDAIGAILSSDOGATWQITPLPFLKGGNMPGR	171
Db	979	SGIEDISVLTGPSTSPETASGVGDLSGLPSGSGLETSASGVF-DVTQLPTERGGLETS	1037
QY	172	GMGERLAVDNNIL-----VFGAPSGKGLWRSTDGATWQMTNFPD-----	215
Db	1038	SGIEDITVLTGRENLETSASGVDEDVSLPSGKEGLETSAGI--EDISVFPTEARGLET	1095
QY	216	--VGTIYAN-PTDTTGYSDIQVWVAFDKSSSLGQASKTIFVGVADPNPNVFWNRDG	272
Db	1096	SASGGVYVSGIPSGEDGTETSTGVEGVS---GLPSGEGLETSASGVEDLGLPTRDSLET	1152
QY	273	GATWQAVPGAPTG-----FIPKHGVDFPNVHLY-----IATNTG-----GPYD-	312
Db	1153	SASGVDTGYPSGREDETSTVP--GVGDDLSGLPSGOEGLETSASGAEDLGLPSGKEDL	1210
QY	313	-GSSGDVWKF-----SVTSGTWTTRISPVPTDANDYFG-----YSGL---	349
Db	1211	VGASGALDGLKPLSGQTPASGLPS-GFSGEYSVDIGSGPSSGLPDPFSLPSG	1269
QY	350	--TIDRHPNTIMVAT-----QISWMPDITIFRSTD--GGATWTRI	386
Db	1270	FPTVSLVDSTLVEVITATTASELEGRGTISVSGSGEESGPPSLSELDSSADISGLPSGTEL	1329
QY	387	WDMTSPNRSRLRVLDISAPFWITFGVQPNP-----PVPSPKLGMW	427
Db	1330	SGOTSGS-----LDVSGTSGFFDVSGQPPGSGSGTGEGTSGIPEVSGQAVRSPDTEI	1382
QY	428	DEAMADPFNSD-----RMLYGTGATLVATNDLTKWDSGGQIHAPMVKGLTEAVNDL	481
Db	1383	SELSSGLSSGQPDVSGEGSLIFGSGSGSITS-----VSG-ETSGISDL	1425
QY	482	ISPPSGAPLISAL-----GDLGFTHADVTAVPSTIFTSPVFTTGTSDVY	526
Db	1426	SGQPSGFPVLVSGTTPGTPLDASGMSGSGDSGTFIVD-----TSLIEVTPITFRE	1476
QY	527	AELNPSIIVRAGSFDPSOPNDRHVAFTDGGKNWQFSGPFGVTTGGTVAASADCSRFX	586
Db	1477	EE-----GLGSVELSGULPSGGETDLSGTSGMVD-VSGQSSGAIDSSGLISPTPEFS---	1525

QY	587	WAPDPCQPVVYAVGFGNSWAASQGVYAN-AQIRSDRVNPKTFYALSNGTFYRSTDG---	642
Db	1526	-----GLPSGVAEVSGEVSGVETGSSLSGAF-----DGSL	1557
QY	643	-----GVTFQVPA--AG-----LPSSGAGVGMFHAVPKGKGLD-----	673
Db	1558	VSGFPTVSLVDRTLVESTITLAPTAQEAEGEPSSILEFSGA---HSGTPDISGDLGSLD	1613
QY	674	-----WLAAS-----SGLYHSTN--GGSSNSAITGVSSAVNVGCKAPGSSYP	715
Db	1614	QSTWQPCWTEASTEPPSPYFSGDFSSTTASGESITAPTG-----SGETSGLP	1662
QY	716	AVFVGT--IGVGTGAYRSDCCGTTWLVINDQHOYGNMGOAITGDHANLRRVYIGTNGR	773
Db	1663	EVTLITSELVEGTEPTVSOELG-----HGPSMT-----YTPLRF-EASGE	1702
QY	774	GIVYDIGG-----APSGSPSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSP	820
Db	1703	ASAGDLGGPVTITFGSGVEASVPEGSSDPSAYPEAGVGVSAAPSAEQSLSEFPLHGIT	1762
QY	821	SSS-----PSP	869
Db	1763	SASRETDMLEMTTPTGVSSNPFTFQEGTRGSAAPVSGESSTSDIDAGTSGVPFATPM	1822
QY	870	SGGVKQYKNNDSAPGDNQIKPGLQVNVNTGSSSVDLSTVT--VRYWFTRDGSGSTLVYNC	927
Db	1823	TSQDRTEISGWSND-----HTEVNVVTVTTPESRWAQSTOHPTELT---	1865
QY	928	DWAAICGNIRASFGSVNPA-----TPTADT	953
Db	1866	-----QEIGSPNPSPSYSGEETQTAET	1885

RESULT 13
VTPX_TTV1
ID VTPX_TTV1 STANDARD; PRT; 360 AA.
AC P19274;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain KRAL) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;
OC Lipothirixvirus.
OX NCBI_TaxID=10480;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174928; PubMed=2308830;
RA Neumann H., Zillig W.;
RT "The TTV1-encoded viral protein TPX: primary structure of the gene and the protein.";
RL Nucleic Acids Res. 18:195-195(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions in no way use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14855; CAA33002.1; -.
DR PIR; S12850; S12850.
KW Repeat.
FT DOMAIN 278 353 3 THR-PRO REPEATS REGIONS AND TWO NEAR IDENTICAL REPEATS.
FT REPEAT 270 291 THR-PRO(N).
FT REPEAT 292 301 THR-PRO(N).
FT REPEAT 302 322 THR-PRO(N).
FT REPEAT 323 332 THR-PRO(N).
FT REPEAT 333 353 THR-PRO(N).
SQ SEQUENCE 360 AA; 38214 MW; EF54FBCB753ED1CA CRC64;

Query Match 4.3%; Score 222.5; DB 1; Length 360;
 Best Local Similarity 22.5%; Pred. No. 0.0025;
 Matches 83; Conservative 55; Mismatches 94; Indels 137; Gaps 17;

QY 581 DGSRFVW-----APGDPGPVVYVAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGT 635
 DB 48 DG---VWVKLRPLSPGOSITI-----TASSGTP-----NIDPTIALYNNNGS 88
 QY 636 FYRSTD-----GGVTFQVAAAG-----LPSSGAVGVMFHAPVCKEG 671
 DB 89 SYSNLVLGSPVTSIVQDEGGYSAISAYASGDFLVASPTGFTPSSRLVW----- 139
 QY 672 DLWL-----AASSGLYHSTNGGWSAITG---VSSAVNVGKGSAPGSSYPAVFVWGT 722
 DB 140 DRWATPTSLDAVGLRLYADTN---DMFGVVRKYINGAQNVSIEQKISG----- 185
 QY 723 IGVGTGAYRSDDCGTWVLND-DQHOYGNWGO-----AITGDHANLRVYIGTNGRI 775
 DB 186 -----YISVNEIDISQFAFTPLVWYLSINGSTANV-KVYQGSNIGT 229
 QY 776 VYGDIGGAPSGSPS-----PSVSPSASPLSPSPSPSSPSPPSPS 815
 DB 230 VSGNYSTPTYPGNPSMAGYGVNDKHYANFIVLPYEPDPQVTVTPISSPSTPTPTPTPT 289
 QY 816 PSSSPS---SSP 861
 DB 290 PTYDITVYVDFVTPSTPTPTLTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 349
 QY 862 PTPSPSPVS 870
 DB 350 PTPSTWSS 358

RESULT 14
 PGCA_MOUSE STANDARD; PRT; 2132 AA.
 AC 061282: Q64021.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN AGC1 OR AGC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C; TISSUE=Cartilage;
 RX MEDLINE=95104847; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fucloep C., Horvath P., Doege K.J., Glant T.T.;
 RT "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan.";
 RT Genomics 22:364-371(1994).
 RN [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.;
 RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene.";
 RT Nat. Genet. 7:154-157(1994).
 CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY

CC -!- SIMILARITY).
 CC DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
 CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
 CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
 CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
 CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
 CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
 CC AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX
 CC DEFICIENCY (CMD). AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY
 CC CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD
 CC CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION
 CC IN THE G1 DOMAIN).
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L07049; AAC37670.1; -;
 CC EMBL: S73722; AAB32160.1; -;
 CC EMBL: S73721; AAB32160.1; JOINED.
 CC HSSP: P98066; ITSG.
 CC MGD: MGI:99602; Agc.
 CC InterPro: IPR003006; Ig_MHC.
 CC InterPro: IPR003596; Ig_v.
 CC InterPro: IPR000538; Link.
 CC InterPro: IPR003324; SGXSG.
 CC InterPro: IPR000436; Sushi_SCR_CCP.
 CC InterPro: IPR001304; lectin_c.
 CC Pfam: PF00059; lectin_c; 1.
 CC Pfam: PF02339; SGXSG; 62.
 CC Pfam: PF00084; sushi; 1.
 CC Pfam: PF00193; Xlink; 4.
 CC ProDom: PD000918; Link; 4.
 CC SMART: SM00032; CCP; 1.
 CC SMART: SM00034; CLECT; 1.
 CC SMART: SM00406; IGV; 1.
 CC SMART: SM00445; LINK; 4.
 CC PROSITE: PS00290; IG_MHC; 1.
 CC PROSITE: PS01241; LINK; 4.
 CC PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 CC PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
 CC Glycoprotein: Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 GI-A.
 FT DOMAIN 152 247 GI-B.
 FT DOMAIN 253 349 GI-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 587 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.

Search completed: July 2, 2002, 09:17:53
Job time: 361 sec

```
QY 129 ANVCMYNSWDNDGAILRSSDQGATWQITPLPKLGGNMPGRGM--GERLA----- 178
Db 2175 ---LFOVRLTTPHDA-----RIRVLPFAFSAGPPTRGLMFGTTLADWRRGKLS 2219
QY 179 -VDPNND--NILYFAPSGKGLWR-----STDGATWSOMTFPDV 216
Db 2220 ETDLPAWRSVPGLTGERGAALGKLSPAQALAAVSVLGRMCLPSTALVALWTCM--FPD- 2276
QY 217 GTYIANPTDTGYSODIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGG--A 274
Db 2277 -----DYTEYDS-----FDALLTARLESQOTL-----SPSGGREA 2306
QY 275 TWQAVPCA---PTGFIPHKGVDFPVNH----- 298
Db 2307 SPPAPPNALYRPTG--OHVAVPAAATHRTPAARVTAMDVLAAVLLGAPVVVALRNTAF 2364
QY 299 -----VLYIATSNITGPP-----YDSSGDVMKFSVTSGTW-----TRISPVSTDTA 340
Db 2365 SRESELELCLTLFDSRARGPDAALRDVAVSDI-----ETWAVRLLHADLNPIENACLA 2417
QY 341 NDYFGYSGLIIDR--QHPNTIMVATQISWMPDTIIFRSTD--GGATWTRIMDWTSYPNRSL 397
Db 2418 AQLPRLSALIAERPLARGPPCLVLVDISMTPTVAVLWENPDPPGPDVRFVG--SEATEEL 2475
QY 398 RYVL---DISA-----EPWLTFCV-----QPNPPVPS 422
Db 2476 PFVAGGEDVLAASATDEDFLARAILGRPFDAALLSGELFPGHPVYQVORAPDDQSPSPNP 2535
QY 423 KLGWMEAMA---IDPNSORMLYGTGAT-----LYATNDLTKWDSGGQIHIA 467
Db 2536 TPGPVDLVGAEGLGPGSLAPTF--TQATGEPVPPRMWAIHGLELELASDDSGGPAPLL 2594
QY 468 ---PMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVETTCTSV 524
Db 2595 APDPLSPTADQSVTSCAPRPGPAVTAREARPGVPAESTRPAP-----VGPRD 2644
QY 525 DYAEINPSIIVRAGSFPDPSOP-----NDRHVAFTDG-----CKNWFOG 564
Db 2645 DFRRL-PS-----POSSAPPDADATAPPPASSASAASSSGSRARRHRRARSILARA 2694
QY 565 SEPGGVITG-----GTVAASADGRFVWAPDGPQPVVYAVGVG-----NSWAASQGV 612
Db 2695 TQASATTQGWPPPALPDTVAPVTDFAR--PPAPKPPPEPAPHALVSGVPLPLGPOAQAQAS 2753
QY 613 PANAQIRSDRVNKTFTYALSNGTYRSTGDGVTFPQVAAAGLPSSGAGVGMFHAVPGKEGD 672
Db 2754 PA---LPIDPVPP-----PVATGTV---LPGCENRRRPLTSGPAPTTPRV---PVGGPQRR 2800
QY 673 LWLAASSGLYHSTNG-GSSW-----SALTGVSSAVNVGFGKSAPGSSYPAVFVVG---T 722
Db 2801 LTRPAVASLSRESLSPMDPADPTAPVLGRNPAEPTSSSPAGSPPPPAVQPVAPPPT 2860
QY 723 IGGVTGAYRSDCCGTWVLINDQHXYGNWQQAITGDHANLRRVYIGTNGRGIVYGDIGG 782
Db 2861 SGPPP-----TYLTLEG-----GV 2874
QY 783 APSG--SPSPSV-SPSASPUSL-----PSPSPSSPSPSPSPSSPSSPSSPS 825
Db 2875 APGGVSRRRPTTRQPVATPTTSARPRGHLTVSRLSAFQPOQPQPOQPQPOQPQPOQPQ 2934
QY 826 PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 885
Db 2935 PQQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 2987
QY 886 D-----NQIKPGLQV-----VNTGSSSDLSLSTVTRYWTFTRDGGSTLVVNCWAA 931
Db 2988 EYPAVRFAQNRPSVPASASSNTRTSGSLSGVSS-----W-----ASSLALHID--- 3033
QY 932 ICGNIRASFSGSVNPATPTADTYLQ 956
Db 3034 -----ATPPPVSLIQ 3043
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 91.42 Seconds
(without alignments)
1810.940 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSNIRLTHRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	47.4	890	2 086727	086727 streptomyc
2	2420.5	47.1	996	2 09AQH0	09aqh0 caldicellul
3	2037	39.7	839	16 097KK0	097kk0 clostridium
4	1784	34.7	806	3 09PAT8	09pat8 agaricus bi
5	1740	33.9	856	3 074170	074170 aspergillus
6	1161	22.6	707	16 09WTE1	09wte1 thermotoga
7	431.5	8.4	1779	2 052374	052374 caldicellul
8	409	8.0	1770	2 09X3P5	09x3p5 caldicellul
9	402.5	7.8	921	2 09L8L8	09l8l8 caldicellul
10	398.5	7.8	1711	2 096311	096311 anaerocellu
11	372.5	7.3	473	10 039620	039620 chlamydomon
12	350.5	6.8	997	2 09Z411	09z411 bacillus sp
13	346.5	6.7	901	2 044562	044562 actinomycetes
14	334	6.5	1751	2 09AQG4	09aqg4 caldicellul
15	325.5	6.3	930	2 09RFX5	09rfx5 caldicellul
16	325	6.3	1000	2 024820	024820 thermophili

17	320.5	6.2	991	10 09AC44	09ac44 chlamydomon
18	318.5	6.2	875	2 09F2B0	09f2b0 thiobacillu
19	308.5	6.0	611	16 074375	074375 synecocyst
20	301.5	5.9	1749	16 073032	073032 synecocyst
21	297	5.8	1915	2 09RPL0	09rpl0 acetivibrio
22	291	5.7	351	10 039492	039492 chlamydomon
23	290.5	5.7	2232	5 091365	091365 caenorhabdi
24	286	5.6	261	2 09AQG7	09aqg7 caldicellul
25	279	5.4	170	2 09RFX6	09rfx6 caldicellul
26	277	5.4	234	12 091GH4	091gh4 epiphyas po
27	276	5.4	1426	2 09X3P6	09x3p6 caldicellul
28	272	5.3	555	10 09FP06	09fp06 chlamydomon
29	265.5	5.2	955	10 094F92	094f92 chlamydomon
30	262	5.1	2468	16 0912M3	0912m3 pseudomonas
31	258.5	5.0	554	6 095L89	095l89 bos taurus
32	257	5.0	1325	5 09BKV7	09bkv7 leishmania
33	254.5	5.0	913	2 091L64	091l64 actinomycetes
34	252.5	4.9	1498	2 09L448	09l448 arthrobacte
35	252.5	4.9	2117	3 096U11	096u11 neurospora
36	247.5	4.8	321	2 046392	046392 clostridium
37	246.5	4.8	818	6 09N1P0	09nlp0 bos taurus
38	245	4.8	887	2 09L3J8	09l3j8 clostridium
39	243.5	4.7	552	10 096343	096343 brassica na
40	243	4.7	288	12 084565	084565 paramesium
41	236	4.6	1003	2 09AHP2	09ahp2 arcanaobacte
42	234.5	4.6	508	2 09F2X3	09f2x3 streptomyc
43	233.5	4.5	3325	12 091BT9	091bt9 turkey herp
44	233	4.5	873	5 09Y076	09y076 leishmania
45	232.5	4.5	687	2 09KVF3	09kyf3 streptomyc

ALIGNMENTS

RESULT 1

086727 ID 086727 PRELIMINARY; PRT; 890 AA.
AC 086727;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE PUTATIVE SECRETED CELLULASE.
GN SC5C7.30C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
MEDLINE=97000351; PubMed=8843436;
RT "A set of ordered cosmid and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031515; CAA20642.1;
DR HSP; P07986; IEXG.
DR InterPro; IPR002860; BNR.
DR InterPro; IPR001919; CBD_2.
DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBD_2; 1.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 47.4%; Score 2436; DB 2; Length 890;
Best Local Similarity 48.6%; Pred. No. 7.4e-115;
Matches 475; Conservative 129; Mismatches 248; Indels 12

Qy	11	MRSRLVSLAATASFAVAALGVPLAITTASPA---HAATTQPYTWSNVAIGGGGVVDGI	68
Db	1	MRRTILITVLLAALAGLLA---GSPPAASAAPAPRAAAADSYTWNARIDGGGVPGI	57
Qy	69	VNEGAPGILYVRTDIGMYRDAANGRWITPLLDWYGNWNGYGVVSTAADPINTNKY	128
Db	58	VFNRTKDLAYRTDIGAYRWEBSHTWTPLLDHWGVGDWGHGTGVVALASDAVDPRV	117
Qy	129	AAVGMVITNSWDNDGAILRSDOGATWQITPLPFKLGNNMPGRGMGERLAVDPNNNDILY	188
Db	118	AAVGTYTNDWPTNGAVLRSADRAGSAWEKADLPFKLGGNNMPGRGMGERLAVDPHDNDVLY	177
Qy	189	FGAPSGKGLWRSYDSGATWSQMTNFPDVGTYIANPITDITGYQSDSIOGVVWVAFDKSS-SS	247
Db	178	LGAIPSGHGLWRSYDAGVTSEVTAFFPNPGNYAODPNDTSGYASDNOGITWVTFDESTGG	237
Qy	248	LGOAKTIFVGVADPNPNPFWSDGGATQWAOVGPATGTFPHKGVPDPVNNHVLVIATSNT	307
Db	238	AGTATRTLYGVADKENAVYRSTDAGATWERLAGOFTGYLAHKGVLDAAENGILYLAYSOT	297
Qy	308	GGPYDGGSGDVWKFVSFTGTWTRISVPSTDTANDFYGYSGLTIDROHPNTIMVATOISW	367
Db	298	GGPYDGGKGLRYATATGTWTDISPAEAADT---YGFSGLTVDQRQCTVMATAYSSW	354
Qy	368	WPDITLFRSDGATWTRINDWTSYPNRSRLRYVLDLSAEPWLTFGVQVPPVPSPKLGMM	427
Db	355	WPDQIFRSDGATWSQAWSYSYSDRENRYTMDVSSPWLTVGANPAPPEQTPKLGMM	414
Qy	428	DEAMADPFNSDRMLXGTGATLATNDLTKW-DSGGQIHIAPMVKGLEETAVNDLISPPS	486
Db	415	TEALEIDPFSDRMVYGTGATVYGTENLTNWDDGGTFAVEPMVGRGLEETAVNDLASPPS	474
Qy	487	GAPLISALGOLGGFTHADVTAVSTTFTSPVFTTGTSDVYAEUNLPSIIVRAGSFDPSSP	546
Db	475	GAPLLSALGVGVGFRHTSLTEVPSMYSPTNFSTTSLDFAETKPDVVVRAGNLD---SGP	532
Qy	547	NDRHVPSTDDGKNWFOGSPBPGVTGGTVAASADGSRFWAPGDPGPVYVAVGFGNSW	606
Db	533	---HIAFSTONGANWFGGTDPSGVSGGTVVAGADGSRFWSP---EGAGVYTTGFGTSW	587
Qy	607	AASQGVPAQAIRSDRNPKTFTYALNSGTFRYRTSDGGVTFQPVAA---GLPSSGAVGVMFHA	665
Db	588	QASTGLPAGALVESDRVNPATFYGFKSGRFPVYSTDGGATFTTAAATGLPAGD---GVRFKA	645
Qy	666	VPKGEDLWLAASS---GLYHISTNGGSSWSAITGVSSAVNVGFKSPAGSSYPAVFV	720
Db	646	LPQGEVDWLGAAGAADPGYGLMHSITDGGGTFTRLPGVDAADTVGFGKAPGASQYLFTS	705
Qy	721	GTIGTVGVARSDDCGTWTWLIINDDOHYGNWGOAITGDHANLRRYVICTNGRGIYVGD	780
Db	706	AETGGVRCIIFRSTDAGATWTRVNDDAHQWGTCAAITGDPVYGRVYVATNGRGIYGD-	764
Qy	781	GGAPSGSPSPSVSPASPSLSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	840
Db	765	-----TSDTGGGTDGPGPDPTP-----	782
Qy	841	SASPSPSSPSSPSSPSSP	900
Db	783	-----TGACEVITYVTNQPWGGFQ---ADVRLTNTGT	811
Qy	901	S-----SVDLS---TVTYRWYFTR--DGGSSPLTVNCDW-----AAIGCENIRA---	939
Db	812	SANNWSLDSFPGGQGEVTRMNAEHTOAGTSVTARNVGNACVAPCASVSGFGTCSRSG	871
Qy	940	-----SFGSVNPATPTA	951
Db	872	TNAEPGCFVAGRACPTA	889

RESULT	Q9AQHO	Q9AQHH	Q9AQHG	Q9AQHF	Q9AQHE	Q9AQHD	Q9AQHC	Q9AQHB	Q9AQHA	Q9AQH9	Q9AQH8	Q9AQH7	Q9AQH6	Q9AQH5	Q9AQH4	Q9AQH3	Q9AQH2	Q9AQH1	Q9AQH0
ID	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
AC	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
CT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
DT	01	02	03	04	05														

Query Match 47.1%; Score 2420.5; DB 2; Length 996;
Best Local Similarity 48.5%; Pred. No. 5.1e-114;
Matches 462; Conservative 151; Mismatches 304; Indels 35;

[illegible]

[illegible]

Qy	794	PSAPSLSPSPSSSPSPSPSPSSPSSPSSPSP 826
		::: : : : : : : : : :
Db	738	ATSTTSVSTTTTTRSTTTTTRSTTTTTS--SGNGSP 769
RESULT 5		
Qy	074170	PRELIMINARY; PRT; 856 AA.
AC	074170;	
DT	01-NOV-1998	(TEMBLrel. 08, Created)
DT	01-NOV-1998	(TEMBLrel. 08, Last sequence update)
DT	01-JUN-2001	(TEMBLrel. 17, Last annotation update)
DE	AVICELASE III.	
GN	AVIII.	
OS	Aspergillus aculeatus.	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.	
OX	NCBL_TaxID=5053;	
[1]		
RP	SEQUENCE FROM N.A.	
RA	Arai M., Takada G., Kawaguchi T., Sumitani J.;	
RT	"Avicelase III from Aspergillus aculeatus.";	
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB015511; BAA29031.1; -	
DR	HSSP; P00725; 2CBH.	
DR	InterPro; IPR002860; BNR.	
DR	InterPro; IPR000254; CBD_fungal.	
DR	Pfam; PF02012; BNR; 7.	
DR	Pfam; PF00734; CBD_1; 1.	
DR	ProDom; PD001821; CBD_fungal; 1.	
DR	SMART; SM00236; fCBD; 1.	
DR	PROSITE; PS00562; CBD_FUNGAL; 1.	
SQ	SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;	
Query Match 33.9%; Score 1740; DB 3; Length 856;		
Best Local Similarity 43.8%; Pred. No. 6.9e-80;		
Matches 366; Conservative 139; Mismatches 297; Indels 40; Gaps		
Qy	30	AALGVLPIATISPAHAATQPTWTSNVAI--GGGGFVDGIVFNEGAPGLIYVRTDIGMY 88
	::: : : : : : : : : :	
Db	4	SSLALLCAALLGLKADAAASQAVTKNVVTTGGGGTTPGIVFNPSAKGVAYARTDIGAY 63
Qy	89	RWDAANGRWIPLLDWVG---WNWNGYGVVSIADPLINTKNVAAVGMVWNSHPDNDGAI 145
	::: : : : : : : : : :	
Db	64	RLN--SDTWTPLMDVGNCDTWHDW---GIDALATDPVTDTRVAVVGMVWNTNEWDPNVGS 119
Qy	146	LRSSDQATGQITPLPKLGNMPGCMGERLAVDPNDNLIYFGAPSGKGLWRSTDGSA 205
	::: : : : : : : : : :	
Db	120	LRSTDQGTWTETKLPFKVGNMPGCMGERLAVDPKNKSILYFGARSHGLMKSTDYGA 179
Qy	206	TWSQMTFPDVGTVIANPTDITGYQSDIOGVVWVAFDKSSSLGQASKIFVGVADPNPN 265
	::: : : : : : : : : :	
Db	180	TWSNVTSEFTGTVYFQDSSST--YTSDPVGIAMVTFDSTSGSGSATPRIFVGVADAGKS 237
Qy	266	VFWSRDGGATQWQAVPGAPT--GFTHPKGVDPDVNHLVLIATSTGGPYDSSGDVWKFVS 324
	::: : : : : : : : : :	
Db	238	VFKSEDAGATWAWVSGEPQYGFPLPKGLVLSPEEKTLIYSYANGAGPYDGTNGTVHKYNT 297
Qy	325	SGTWTRISPVSTDTANDFYGYSGLTIDRHPNTINVATOISWMPDTIIFRSTDGGATWT 384
	::: : : : : : : : : :	
Db	298	SGVWTDISP---TSLASTIYGYGGLSVLDLQVPTLMVAALNCWNPDELIFRSTDGSA 354
Qy	385	RIWDWTSPNRSRLVLDISAEPLWLTGVPQNP--PVSPKLGWMDEAMAIIDPFNSDRMLY 443
	::: : : : : : : : : :	
Db	355	PIWENWGYPISINYYSDISNAPWIODTTSTDDQFPV---RVGWNVEALAIIDPPDSNHL 411
Qy	444	GTGATLTATNDLTKWDSGGQIHAPVWVGLFEETAVNDLISPPSCAPLISALDGLGFT 503
	::: : : : : : : : : :	
Db	412	GTGLTVYGGHDLTNWDSKHNVTVKSLAVGATEEMAVLGLITPPGGPALLSAVGDGDFYHS 471
Qy	504	DVTAVPSTIFTSPVFTTGTSDVDAELNPSIIVRAGSFDSPSQNDRHVAFSTDGGKNWQF 563


```
Db 472 DLDRAPOAHTPIYGTNGIDYAGNKPNSIVIRSGASD--DYPT--LALSSNFGSTWYA 526
QY 564 GSEPGGVTTGTTVAASAGSRFYWAPDGPQVYVYVGFNSWAAASOGVAPANAQIRSDRV 623
Db 527 DYAASTSTGTGAVALSADGTVLLMSSTSGALVSKSQG---TLTAVSSLSFGAVIASDKS 583
QY 624 NPKTFYALSNGTFRYRTDGGVTFQPVYAAGLPSSGAVGVMFHVPKGGEDLWLAASSGLYH 683
Db 584 DNTVFYGGAGAIYVSKNTATSFKTYS-LGSSTTVNAI-RAHPSTIAGDVWASTDKGLWH 641
QY 684 STNGSSWSAI-TGVSAVNVGFGKSAPGSSYPVAVFVVGITGGVGTGAYRSDDCGTTWVLI 742
Db 642 STDVGSFTIGSGVTAGNFGFGKASSTGCVYVYGFYFFIDGAAGLFKSEDACTNMQVI 701
QY 743 NDDHQYGNMQA--ITGDHANLRVYIGTNGRIVYDGGAPSG-----SPSPSV 792
Db 702 SDASHGFGS-GSANVNVGDLQTYGRVFRGHERPGHLLRQSPREPAGRHGDDGDTTTSKT 760
QY 793 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852
Db 761 STVSTTLKTTSSASTSSSTVTKTTSSSTTSKASSTTTTNTTTTSSGTTATA 820
QY 853 SS 854
Db 821 SA 822

RESULT 6
ID OSWYEL PRELIMINARY; PRT; 707 AA.
AC OSWYEL;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOGLUCANASE, PUTATIVE.
OS TM0305.
GN Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001712; AAD35393.1;
DR TIGR; TM0305;
DR InterPro; IPR002860; BNR.
DR Pfam; PF02012; BNR; 9.
KW Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CBDA6A99F CRC64;
```

Query Match 22.6%; Score 1161; DB 16; Length 707;
Best Local Similarity 34.6%; Pred. No. 7.2e-51;
Matches 267; Conservative 126; Mismatches 280; Indels 98; Gaps 27;

```
QY 32 LGVLPIAITASPAHAANTQPTWNSVNAIGGGFVDGIVNEGAPGILYVTRDIGMYRWD 91
Db 10 LAIGLVVFGAT-----FEWKSVEINGGVFPGIIFHPASPGLLYARDVGGLYRWD 60
QY 92 AANGRWIPLLDWGNWNGYGVVSIADPTNTNKVAAVGMVNTSDPDNDGAILRSSDQ 151
Db 61 EETRWKQLDFLRDQSDYGVLSVALDSPDKRIYATGKYTDW-AGYGAILISEDY 119
```

```
QY 152 GATWQITPLP---FKLGGNMPGRGMRGLAYDPNNDNILYFGAPSGKGLHRSDSGATWS 208
Db 120 GETTIVNLNDKYIKVGGNEDGRNAGERLQVDPNFSSVLEMG-T-KYGLWKSEDFGNWK 178
QY 209 QMTNFPDVGVIYIANTDTTCYQSDIQGVVWAFDKSSSSSLGQASKTIFVGVADPNPNVFW 268
Db 179 KVDSPFST-----SVTFVLFDEKSGEKGSPTPRIFVGCSPKG-IFV 219
QY 269 SRDGGATQWAPGAPTGFIPHKGVDFPNVHVLVIATSNGTGGPYDGGSSGDKWYKSVTSGTW 328
Db 220 TEDGGTTNVNLPNLDIPLRGKIH--DGILYVTLNAGLPGNGATRGAVKVVYADQKW 277
QY 329 TRISPPVSTDTANDYFGYSLGTLTDROHPNIMVATQISWHPDIIIPRSTDGGATWTWID 388
Db 278 YDTYPMKGD-----FGYCGIDVQE---NVVISTLDRWYPHDEIFISLNGGETWTRPLE 328
QY 389 WTSYPNRSRLRYLVDISAEPWLTFCGVQNPVPSPKLGWMDMAIDPFNSDRMLYGTGAT 448
Db 329 KANF-----DINKAPWK---DLNP-----HWISD-VKIDPFDMNRAIFTTGYG 368
QY 449 LYATNLTWKDSGGQIHIAPMVK-----GLEETAVNDLISPPSGAPLISALDGLGFTW 502
Db 369 VWTYELKKSFG-----MGRPKVWKFENRGLLEETVVLQLPPIGERPLLSAIADWGGFRH 424
QY 503 ADVTAVPESTIFTSPVFTTGTSDYAEINPSIIVRAGSFDPSOPNDRHVAFTDGGKNWF 562
Db 425 ESLDTPSSMY-KPLKWTSLGIAFAYONSKFVARVHTY---TYP---FLSISEDGGINWR 477
QY 563 Q-GSEPGGVTTGG--TVAASADGSRFYWAPDGPQVYVYVGFNSWAAASOGVPA---NA 616
Db 478 EIEVPEGITDGGRLSLAVSNDGKTLWSPAN--HEVIVSSDKGSKWKKAISVPVPEFY 535
QY 617 QIRSDRVNPKTFYAL--SNGTFFRSTDGGVTFQPVYAAGLPSSGAVGVNFHA----VPGKE 670
Db 536 FPASDPVNPSEFYIFDKWNGDLFLSKDGGSEFK-KAKLPSEFDNWYSLSYFPLADPRE 594
QY 671 GDLWLAAS-SGLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPVAVFVGTIGGYTGA 729
Db 595 GDIWLALQWNLGRSKDGGITFERLGNVDIAYVIGFAPKPGTDPYPAIYLNGVNGVYGI 654
QY 730 YRSDCCGTTWVLINDDQHQYGNMQAITGDHANLRVYIGTNGRIVYIGDI 780
Db 655 FMSTDEGKTWNRINNDKHQFG-WIHYMIGDMNEGRIFLTGEGRIIVGEV 704

RESULT 7
ID O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.8).
GN XVNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1.
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036924; AAB95326.1;
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBD_3; 3.
DR Pfam; PF02018; CBD_6; 2.
DR Pfam; PF00331; Glyco_hydro_10; 1.
```


[illegible]


```
RESULT 10
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)
DE (CELLULOSE A) (FRAGMENT).
GN CELA.
OS Anaerococcus thermophilum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Anaerococcus group; Anaerococcus.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A..
RC STRAIN=2-1320;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E
CC (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L
CC (FAMILY 48 OF GLYCOSYL HYDROLASES).
DR EMBL; Z86105; CAB06786.1; -.
DR HSP; P26221; ITF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR Pfam; PF00942; CBD_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRASE48.
DR ProDom; PD001947; CBD_3; 2.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL-HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL-HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER 1
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 7.8%; Score 398.5; DB 2; Length 1711;
Best Local Similarity 22.9%; Pred. No. 4.3e-12;
Matches 197; Conservative 116; Mismatches 267; Indels 281; Gaps 40;

QY 262 PNPVFNRSRGATWQA-----VPGAPGFTPHKGVDPVNVHLVIATSTNGGPDGS--S 315
DB 25 PN-----NWRGDSALKDQDGLDLTGWFADGDKVXF-----NLPMSYTGTFMS 73
QY 316 GDVWKFSS---VTSGTWTRISVPSTDFANDYFGVSGULTIDROHPNTIMVATQI----- 365
DB 74 WAVEYKDAFAVKSQLEHI--LNQIEWNDYF-----VKHPKSVKYYVQVGDGSKDH 124
QY 366 SWW-----PDTIFRSTGGATWTRTWMTWTSYPNRSLRYLDISAEPMLTFGVQP 415
DB 125 AWGPAEVMQWERPSFKVTQSSPGSTVVTE---TAASLAASAIVLK-----DR 169
QY 416 NPPVPSPKLGWDEMAIDFNSDRMLYGTATLYATNDLTWKDSCGQIHIAWPVGLLEE 475
DB 170 NPKAATYLLQHAKE-----LYEFAEVTKSDAG-----Y 197
QY 476 TAVNDLISPPSG-----APLISALDGLGGFTHADVTAVPSTFTSPVETGTGSDVY 526
DB 198 TAANGYNSGSGFYDELSSAAVWLYLATNDSTYLTAE-----SYQVONWPKISGNTIDY 252
```

```
QY 527 -----AELNPISIIVRAGSFDPSQPNDRHVAFTDG-----GKNWFO 563
DB 253 KWAHCWDDVHNGAAL---LLAKITGKDIYKQIIESHLDIYWTGYNGERIKYTPKGLAWLD 309
QY 564 --GSEPGGVTTGGTVAASAD-----GSRFVWAPGDPGPVYVAVGFG-- 603
DB 310 QWGLRYATTATFLAFVYSDWVGCPSTKKEIYRKFGESQIDYALGSAGRS--FVVGFGTN 367
QY 604 -----NSWAASQGVPA-----NAQIRSD--- 621
DB 368 PPKRPHRHTAHSSWADSQSIPSYHRRHTLYGALVGGPGSDSDSYTDDISNYVNVNACDYNA 427
QY 622 -----RVNP-----KTFYALNSGTFFYRSTDGGVTFQFVAAAGLPSGAGVGMFH 664
DB 428 GFVICALAKMYQLXGGNPIDPFKALETNTDEFF-----VEAGINASGTNFEIK 476
QY 665 AVPGKEGDLWLAASS-----GLYHSTNGSSWSAITGVSASVNVWGKKSAP----- 710
DB 477 AIYNNQSG-WPAKATDKLKFYFVDSLSELIKAGYSPNQLT--LSTNYNOGAKVSGPYWMDA 534
QY 711 --GSSYPAVFVVGTI--GGVTGAYRSDDCGTTWVLINDQHQYGNWGOAITGDHANLRRV 766
DB 535 SKNIYILVDFGTGLYPGGQDKYKE-----VQFRIAPQNVQWVNDNSNDYSFODIKGV 588
QY 767 YIGT-----NGRGIYVGDIGGAPSGSPSPSVSPASPSLSLSPSPSPSPSPSPSPSP 816
DB 589 SSGSVVTKYIPLYDGDVKVMDGPGGSGATPTTA-----TATPTTPTVTPPTPTPTP 642
QY 817 SSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 875
DB 643 TS--TATPTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 690
QY 876 QYKNDSAPGDNQIKGLQVNVNTGSSSVDLSTVTYRVWFTRDGGSSFLVNCDDAAIGCG 935
DB 691 LYANKETNSTNTIRPWLKVNTGSSSIDLSRVTIRYWTVDGDKAQSALS-DWAQIGAS 749
QY 936 NIRASFGSVNPATPTADTYLQ 956
DB 750 NVTFKFKVLSVSGADYYLE 770

RESULT 11
Q39620 PRELIMINARY; PRT; 473 AA.
AC Q39620;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VSP-3 PROTEIN PRECURSOR.
GN VSP-3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC-621;
RC MEDLINE=95093034; PubMed=8000007;
RA Woessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,
RA Goodenough U.W., Haring M.A.;
RA "Domain conservation in several volvoclean cell wall proteins.";
RL Plant Mol. Biol. 26:947-960(1994).
DR EMBL; L29029; AAB53953.1; -.
KW Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 473 POTENTIAL.
SQ SEQUENCE 473 AA; 47532 MW; 38506131FAA674A4 CRC64;

Query Match 7.3%; Score 372.5; DB 10; Length 473;
Best Local Similarity 33.0%; Pred. No. 2e-11;
Matches 132; Conservative 36; Mismatches 139; Indels 93; Gaps 17;
```



```
Db 241 TSGFIDDLGW---AAVWLXIATNDSYLTAKAELMSEYANGTNTWTQCWDVRYGLIM 297
QY 100 LLDVWGNWNGYNGWSTADPINTNWAAGVM-YTNSWDPNDAIRSSDQG----- 152
Db 298 LAKITGREL--YKCAVERNLDHWDTRITYPKMAYLTGW---GSLRYATTAFLACVY 351
QY 153 ATWQ-----ITPLPKLGNMPGR-----GMGERLAVDPNNNNILYFGAPSG 194
Db 352 ADWSGCDNKKTKYLNFAKSIDYALGST--GRSFVVGFGTNYPOHPHRR----- 400
QY 195 KGLWRSTDGATWSOMTFDP-----VGYIANPTDTTGYQSDIOGVVW--VAFDKSS 246
Db 401 -----AHSSWANSWKIPEYHRIHLYGALVGGPGSDSDTDITDYONEVACDYNAG 452
QY 247 SLGQASTIFGVADPNPNVFWSDGGATWQAVPGATGFIHPKGVDPVNVHLYIAT-- 304
Db 453 IVGALAK-----MYQLYGEPID--DFKAIETPTNDEIFVESKF 489
QY 305 SNTGCPYDGGSGDVWKFVTSVGTWTRISVPSTDTAN-DYF-----GYSGLTTDRQ 354
Db 490 GNSQGP---NYTEVISIYIYRTGW---PPRVTDKLSFKYFIDLTETLIQAGYS----- 535
QY 355 HPNTIMVATOISWPDITIIERSTDGGATWTRWDWTSYPNRSRLRYVL-DISAEPLTFGV 413
Db 536 -PDVVKVDT-----YIEGKLSGPYVWD-----KNRIYIYLVDSGTK-----I 575
QY 414 QPNPVPSPKLGW---MDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM 469
Db 576 YPGGEVEHKKQAQFKISVPGQYWDPTN-DPSYKGLTSQLEKNKYIAAYDNN-----L 628
QY 470 VKGLEEATVNDLISP-----PSGAPLISALGDLGFGTHADVTAVPSTFT-----SPVFT 519
Db 629 VWGLEPGAATSTPATPTPTPTPTPTVTA-----TPPTPTPTPTGSP--G 674
QY 520 TGTSDVYAEINPSIIVRAGSFPDSSQPNDRHVAFSTDGKNWFOGSEPGVTTGCTVAAS 579
Db 675 TSGVKVLYKNNETSASTGSIKP-----WFK-----IVNGG--SSS 708
QY 580 ADGSRF---VW--APGDPGPVYVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSG 634
Db 709 VDLRVRKIRYWTYVDGDKPOSVC-----DW-----AQIGASNVT--FNEVKLSSG 752
QY 635 TFYRSTGGVTFQVPAAGLPSGAGVGNFHAVPCKE--CDLWLAASSGLYHSTNGSSWSA 693
Db 753 V-----SGADIY--LEVGF--SSGAGQLQ-----PKDGTGDIQVRFNKNDWSYNOADDWSW 800
QY 694 ITGVSSAVNVGFGKSPAGSSYPVAVVVGTTGGVTGAYRSDDCGTTWVLINDDQHQYGNWG 753
Db 801 LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG 824
QY 754 QAITGDHANLRVYIGTNGRGIVYGDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPS 813
Db 825 Q-----EPGGA-----TPAPTSTATPT 841
QY 814 PSPSSSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 873
Db 842 PTPATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 887
QY 874 KVQKNNDSAPGDNQIKPGLQVWNTGSSVDLSVTVRYWFTDRDGGSTLVNCDWAAIG 933
Db 888 KWWANGNLSPNTVNLNPKIKIENVGTAVDLSRVKRYWYTIIDGATQSV----- 938
QY 934 CGNTRASFSGSNPA 947
Db 939 -----SVASSINPA 947
```

RESULT 15

Q9RFX5

ID Q9RFX5

AC Q9RFX5;

DT 01-MAY-2000 (Tremblrel. 13, Created)

PRT; 930 AA.

```
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
MULTIDOMAIN BETA-1.4-MANNANASE PRECURSOR.
GN MANA.
OS Caldicoccus cellulosivorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10633733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicoccus cellulosivorans and action of the recombinant enzyme on
RT kraft pulp."
RL Appl. Environ. Microbiol. 66:664-670(2000).
RE EMBL: AF163837; AAF22274.1; +.
DR HSP: O06851; INEC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR004302; Chitin_bind_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF00942; CBD_3; 2.
DR Pfam: PF00150; cellulase; 1.
DR PRINTS: PR03067; Chitin_bind_3; 1.
DR PRODOM: PD001947; CBD_3; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT CHAIN 34 930 MULTIDOMAIN BETA-1.4-MANNANASE.
FT SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;
SQ
Query Match 6.3%; Score 325.5; DB 2; Length 930;
Best Local Similarity 25.8%; Pred. No. 9.9e-09;
Matches 109; Conservative 59; Mismatches 98; Indels 157; Gaps 21;
QY 560 NWF-----OGSE-----PGVTTGTVAAASADG---SRFW--APGDPGPVYVAVGFG 603
Db 81 NWFENLISNAGRRHREIIPDGKLGPT--ASFDMNQARTDMMWTRLOPGATITVRV--- 135
QY 604 NSWAASQGVPAANAQIRSDRVNPKTFYALNSCTGYRSTGGVTFQVPAAGLPSGAGVME 663
Db 136 NAWAP-----HPGTWY-----LYVTRDGDWDTQPL----- 160
QY 664 HAVPGKEEDLWLAASSGLYHSTNGSSWSA--TGVSANVNGFGKSP--GSSYPVAVV 720
Db 161 -----KWSLEPTPFQVTPNPPINSSGPDGAEY----- 188
QY 721 GTGGVTCAYRSDDCGTW-VLINDDQHQ---YGNWQAITGD-HANLRVYIGTNGRGI 775
Db 189 -----SWQVLFNKGRIIYIMWRSDSPEAFYNCSDYFGSGPIAY 231
QY 776 VYGD--IGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 833
Db 232 EFGDPREGGT-----MITP-----PPSGTPTPTPTPTPTPTPTPTPTPTPTPT 273
QY 834 PSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 893
Db 274 P-----IVTPTSTPTPSA-----SGLRVEYRVGDTSATDNQMKPQL 310
QY 894 QVWNTGSSVDLSVTVRYWFTDRDGGSTLVNCDWAAIGCGNTRASFSGSNPATPTADT 953
Db 311 RIVNTGSAVPLTELKVRVWYTKN-STQAEQYFCDNAIGCSNIRAOFSVLSQPVSGADS 369
QY 954 YLQ 956
Db 370 YIE 372
```

Search completed: July 2, 2002, 09:17:22

Job time: 330 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:34 ; Search time 3473.69 Seconds
(without alignments)
17283.705 Million cell updates/sec

Title: US-09-917-376-2
Perfect score: 2869
Sequence: 1 atgtagcttcggagaacat.....ggcggacacctactctgcagn 2869

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

C	1	748	26.1	41906	1	SC5C7	AL031515 Streptomy
	2	512	17.8	3262	1	AF078038S1	AF078038 Caldicell
	3	402	14.0	2745	8	AB015511	AB015511 Aspergill
	4	375.6	13.1	12732	1	AE007608	AE007608 Clostridi
	5	327.6	11.4	2579	8	AB1292929	AJ292929 Agaricus
	6	131.6	4.6	3237	1	AF200304	AF200304 Caldibaci
	7	103.6	3.6	749	14	NFORP2	D13755 Multicapsid
	8	103.6	3.6	131995	14	OPU75930	U75930 Orgyia pseu
	9	100.8	3.5	390	6	AR135147	AR135147 Sequence
	10	96.4	3.4	4567	1	AF163837	AF163837 Caldibaci
	11	91.4	3.2	183665	2	AL645600	AL645600 Mus muscu
	12	90.2	3.1	40340	1	AP003605	AP003605 Nostoc sp
	13	86.4	3.0	183544	2	AC095146	AC095146 Rattus no
	14	84.2	2.9	77225	2	AC099417	AC099417 Rattus no
	15	83.6	2.9	118584	14	AY043265	AY043265 Epiphyas
	16	82.2	2.9	180104	9	DJ526N18	AF123462 Homo sapi
	17	81.6	2.8	13669	14	TTV1TP	X14855 Thermoprote
	18	81.6	2.8	99427	2	AP003821	AP003821 Oryza sat
	19	81.2	2.8	216819	2	AL645725	AL645725 Mus muscu
	20	80	2.8	5437	1	CASR69XYN2	AF036924 Caldicell
	21	80	2.8	11707	1	AF078737	AF078737 Caldicell
	22	80	2.8	11707	6	E35099	E35099 Truncated c
	23	79	2.8	15348	1	AE007163	AE007163 Mycobacte
	24	78.4	2.7	159974	2	AC021474	AC021474 Homo sapi
	25	78.2	2.7	77457	1	AF210249	AF210249 Streptomy
	26	78	2.7	1256	7	AF069776	AF069776 Xanthomon
	27	77.6	2.7	16671	1	AE006990	AE006990 Mycobacte
	28	77.6	2.7	67200	1	MTV017	AL021897 Mycobacte
	29	77.4	2.7	125020	9	AF429315	AF429315 Homo sapi
	30	77.2	2.7	63621	9	AL590867	AL590867 Human DNA
	31	76.8	2.7	4034	14	TTV1VP	X14717 Thermoprote
	32	76.8	2.7	165742	2	AC020697	AC020697 Homo sapi
	33	76	2.6	169915	2	AC068801	AC068801 Homo sapi
	34	76	2.6	173434	2	AC016142	AC016142 Homo sapi
	35	76	2.6	176463	2	AC012149	AC012149 Homo sapi
	36	75.8	2.6	13278	1	AE001712	AE001712 Thermotog
	37	75.6	2.6	203299	2	AP001178	AP001178 Homo sapi
	38	75	2.6	5513	1	AT286105	Z86105 A. thermophi
	39	74.8	2.6	47852	1	MTV023	AL022022 Mycobacte
	40	74.4	2.6	1171	14	BPU10884	U10884 Bacterioph
	41	74.4	2.6	15311	1	AE007164	AE007164 Mycobacte
	42	74.2	2.6	161355	2	AC011154	AC011154 Homo sapi
	43	74.2	2.6	192931	2	AC094068	AC094068 Rattus no
	44	74	2.6	11186	1	AE003939	AE003939 Xylella f
	45	73.6	2.6	56414	1	MTV002	AL008967 Mycobacte

ALIGNMENTS

RESULT	1	SC5C7	41906 bp	DNA	linear	BCT 07-SEP-1998
LOCUS	SC5C7	Streptomyces coelicolor cosmid 5C7.				
DEFINITION	AL031515	Streptomyces coelicolor protein; lyase; narg; nargH; nari; narJ; nitrate reductase; oxidoreductase; RNA polymerase sigma factor; secreted cellulase; uvrA-like protein.				
ACCESSION	AL031515.1	GI:3559990				
VERSION	AL031515.1	GI:3559990				
KEYWORDS		Streptomyces coelicolor A3(2).				
SOURCE		Streptomyces coelicolor A3(2).				
ORGANISM		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.				
REFERENCE	1	(bases 1 to 41906)				
AUTHORS		Seeger,K.J. and Harris,D.				
JOURNAL		Unpublished				
REFERENCE	2	(bases 1 to 41906)				
AUTHORS		Parkhill,J., Barrell,B.G. and Rajandream,M.A.				
JOURNAL		Direct Submission				
TITLE		Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project,				
JOURNAL		Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge				
		CB10 1SA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Prof.				
		David A. Hopwood, [3] John innes Centre, Norwich Research Park,				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Colney, Norwich, Norfolk NR4 7UH, UK
3 (bases 1 to 41306)
Redenbach,M., Kieser,H.M., Denapaita,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.

Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.

(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are

numbered using the following system eg SC7B7.01c. SC (S.

coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene

prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program

of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/>

Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtc, ttg or (att)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most

upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the

sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 5C7 lies

between 1E6 and 4B5 on the AseI-A genomic restriction fragment.

Location/Qualifiers

FEATURES
source

1..41306
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid 5C7"

complement(1..136)

/gene="SC5C7.01c"

complement(<1..136)

/gene="SC5C7.01c"

/note="SC5C7.01c, unknown, partial CDS, len: >45 aa"

/codon_start=1

/transl_table=11

/product="hypothetical protein SC5C7.01c"

/protein_id="CAA20613.1"

/db_xref="GI:3559991"

/db_xref="SPTREMBL:O86698"

/translation="MQRGSDRMVHRDDEMKHELOGLLRSGHPTRSEWNDEPAEAD
D"

complement(142..146)

/note="possible RBS upstream of SC5C7.01c"

336..339

/note="possible RBS upstream of SC5C7.02"

345..2735

/gene="SC5C7.02"

345..2735

/note="SC5C7.02"

to Streptomyces peucetius TR:Q54827 (EMBL:L76359)
daunorubicin resistance protein drrc (764 aa), fasta

scores: opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity
in 775 aa overlap, and to many uvra proteins e.g.

UVRA_NEIGO excinuclease ABC subunit A (950 aa), fasta

scores: opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in
953 aa overlap. Contains 2x PS00017 ATP/GTP-binding site
motif A (P-loop) and 2x pfam match to entry PF00005
ABC_tran, ABC transporters, scores 42.40, E-value 1.7e-10
and 71.20, E-value 2.2e-17"

/codon_start=1

/transl_table=11

/product="uvrA-like protein"

/protein_id="CAA20614.1"

/db_xref="GI:3559992"

/db_xref="SPTREMBL:O86699"

/translation="MSSAKRPGTGPESHVADSHDLIRVHGARENKLDVSDIPKRR
LVFTGVSQSGKSLVFNITAAESQRLINETYSAFVGQFMPTLARPEVDVGLTAT
IVDQRMGADPRSTVGTATDVNAMRLILFSLRGEPRIGPPSAYSENTASVRASGATV
ERGNKAVRATERTGGCTHCEGRGTVSDIDLTYLDDSKSLAGGAFIPGWKSDSQ
WTVQVYQSGFVDPDKPIREYTEKRLDELYGEPVKVNGVNLVYEGILPKIQSEFL
SDKRAMQPHIRAFVERAVFTTCPECEGTRLSEGRSKIKKSIADACAMEIDRLA
EWRDLTSPVAPLTALTALDLSFEIGLYLSLDRPAGTUSGGEAGAKVMIRHLGS
STDTYTFDEPTVGLHDIQRMNDLLRLKNGTVLVVHKPEATAIADHVVDLG
PGAGTAGTVCFTGVEELRAADVTGRHLDRAVLKESVRPAGALETRDARTNLQ
GVDVPLGLCVKVTGAGSGKSLIHGSPAGADVYTDQSPKIGSRSSNPATVTGL
LDPIKAPAKANGVKPALFSANEGACPTCNGAGVYTDLMAGVATPCDECEGRF
QPAYLEFRGGRDISEVLAMSYDQAEFFGAGEARTPAHKILQRLSDVGLYLILGQ
PLTUSGGERQRLKATHMGKGVTVLDEPTTGLHLDVDEQLGLLDRVQAGSVI
VIEHHQAVMAHADWILDLGPGAGHDGGRVVFEGPADLVADRSTUTGHELAAYGA"

471..1700

/gene="SC5C7.02"

/note="Pfam match to entry PF00005 ABC_tran, ABC

transporters, score 42.40, E-value 1.7e-10"

492..515

/gene="SC5C7.02"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

1893..2627

/gene="SC5C7.02"

/note="Pfam match to entry PF00005 ABC_tran, ABC

transporters, score 71.20, E-value 2.2e-17"

1914..1937

/gene="SC5C7.02"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

2803..3285

/gene="SC5C7.03"

2803..3285

/gene="SC5C7.03"

/note="SC5C7.03, probable transcriptional regulator, len:

160 aa; similar in N-terminus to many e.g. RAMA_ENTCL

transcriptional activator rama (113 aa), fasta scores;

opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96

aa overlap. C-terminal 50 aa are unique. Contains pfam

match to entry PF00165 HTH_2, Bacterial regulatory

helix-turn-helix proteins, arac family, score 71.40,

E-value 1.9e-17"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator"

/protein_id="CAA20615.1"

/db_xref="GI:3559993"

/db_xref="SPTREMBL:O86700"

/translation="MTTLEDLARLRRLDMREYAEPDLPALARVALMSAGHSRS
FRAAYGEPTYSYLMTTRRIERIMALLRRGDMSTVCFAVGCTSLGFSRFTLVGES
PSAYRERDHDGAALPACVAKVYVTRPVRNGEGRSGTAPAGKSESVRNGEAPAPPPVA
"

2872..3132

/gene="SC5C7.03"

/note="Pfam match to entry PF00165 HTH_2, Bacterial

regulatory helix-turn-helix proteins, arac family, score

71.40, E-value 1.9e-17"

3291..3707

/gene="SC5C7.04"

3291..3707

/gene="SC5C7.04"

/note="SC5C7.04, possible lyase, similar to e.g.

LGUL_HAEIN lactoylglutathione lyase (EC 4.4.1.5) (135 aa),

fasta scores; opt: 124 z-score: 215.2 E(): 9.7e-05, 29.0%

misc_feature

gene

CDS

RBS

RBS

gene

CDS


```

identity in 131 aa overlap"
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAA20616.1"
/db_xref="GI:3559994"
/db_xref="SPT:EMBL:O86701"
/translation="MDISLSOCFTAVDDHDLALAFYRDVLGLEVRNDVGVEGRWVTL
GSPAQPDVIVLEPLDPAKADQAMADLLAKLHGLGIILATDDVDATFCHVRRAA
GAELVQEPDQPVGRCAFRDPAGNVMVRLHRPKG"
3978..3981
/note="possible RBS upstream of SC5C7.05"
3994..4794
/gene="SC5C7.05"
3994..4794
/gene="SC5C7.05"
/note="SC5C7.05, probable RNA polymerase sigma factor,
len: 266 aa; similar to S. coelicolor sigma factors of the
sigB subfamily e.g. TR:052313 (EMBL:AF036131) putative RNA
polymerase secondary sigma factor sigQ (263 aa), fasta
scores: opt: 1258 z-score: 1540.5 E(): 0. 72.9% identity
in 251 aa overlap, and to RPOF_STRAU RNA polymerase
sigma-F factor (297 aa), fasta scores: opt: 713 z-score:
858.3 E(): 0. 42.6% identity in 312 aa overlap. Possible
alternative start 216 bp upstream, not supported by
FramePlot. Contains helix-turn-helix motif at aa 228-249
(score 2104, +6.35 SD) and Pfam match to entry PF00140
(sigma70. Sigma-70 factors, score 210.40, E-value 2.8e-59"
/codon_start=1
/transl_table=11
/product="putative RNA polymerase sigma factor"
/protein_id="CAA20617.1"
/db_xref="GI:3559995"

```

Query Match	26.1%	Score 748;	DB 1;	Length 41906;
Best Local Similarity	59.8%	Pred. No. 1.4e-74;		
Matches 1429;	Conservative 0;	Mismatches 895;	Indels 64;	Gaps 8;
QY 154	tacacctggagcaacgtggcgatcggggggcgcgctttgtcagcggatcgtcttcaat	213		
Db 34351	TACACCTGGAGAACGCCGCATCGACGGCGCGGTTCGTCCCGCATCGTCTTCAAC	34292		
QY 214	gaagggtcacccgggaattctgtactgtcgagcagacatcgggggagtgtatcgatggat	273		
Db 34291	CGGACCGAGAAGACCTGGCTACGCCCGCACCGACATCGAGCGCGCTACCGCTGGCAG	34232		
QY 274	gcgcgcaacgggggtggatccctctcttgattggattgggtggaacaattgggggtac	333		
Db 34231	GAGGAGTCCACACACTGGAGCGCGCTCTCGACACACTCTGGCTGGGACGACTGGGGGCAC	34172		
QY 334	aacggcgctgcagcatctggcgagacccgatcaataactaacaaggatggcgccgcgtc	393		
Db 34171	ACGGCGTGTGCTCTCGCTCGACGCGCTCGATCCGAGCCGGGTACCGCGCGCGTC	34112		
QY 394	ggaatgtacaccaacagctgggacccaaaacgacggagcattctccgtcgtctgtatcag	453		
Db 34111	GGCAGGTACACCAACGACTGGGACCGCACCAAGCGCGGTGCTCGCTCGCGCACCGG	34052		
QY 454	ggcgcaacgtggcaaaataacgcccctcggttcaagcttggcggaacatgcccgggcgt	513		
Db 34051	GGTCCGAGCTGGGAGAAAGCGGACACTCCGCTTCAAGTGGCGGCAACATGCCCGCGCGC	33992		
QY 514	ggaatggcgagcgcttgcggtggatccaaacaatgacaacatctgtatttcggcgcc	573		
Db 33991	GGCATGGGTGAGCGCCTGGCGCTGCACCGCACACGACGTGCTGTACCTGGGCGCC	33932		
QY 574	ccgagggcaaaaggctcttggaagcacagatcccgcgcgacctggtccagatgacg	633		
Db 33931	CCCAGCGCCACGAGACTGTGGCGGTCCACCGACGCGGGCGTCACTTGTGGAGGTGACG	33872		
QY 634	aactttccgagctaggcaagtcacattgcaaatcccaactgacacgacgggctatcagac	693		
Db 33871	GCCTCCCGAACCCGGGAACACTCGCGAGGACCCGACACGACGTCGGGTACGCGCTCC	33812		

Qy	694	gatattcaaggcgtctgctgtgggtcgcttttcagaaagtcttgcgtcatcgtcgtcggaacg	753
Db	33811	GACAAACAGGGCATACCTCTGGGTACCTTCAGAGTACGACGGCGCGCGGGCAGC	33752
Qy	754	agtaa--gacatttttggtggcgtggcgatcccaataatcccggtctcttggagcaga	810
Db	33751	GCCAGCGGACCCCTCTACGTCGGGTGCGGACAGGAGACGCGCTCTACCGCTCGACC	33692
Qy	811	gacggcgcgagcgtggcagggcggtgcggggtgcgcgcgcgcgttcatcccgcaag	870
Db	33691	GACGCGGCGCGACCTGGGAGCGGCTCGCGGGACAGCGACGCGGTACTGTGCCACACAG	33632
Qy	871	ggcgtcttgaccgggtcaaccagcgtcctctatatgtccaccagcaatcacgggtggtccg	930
Db	33631	GGTGTGTCGACGGGAGAACGGGTACCTGTACCTTCGCTACAGGCACACCGCGGCCCG	33572
Qy	931	tatgacgggagctcggcgacgtctggaaattctcgtgacctcgggacatggacgcga	990
Db	33571	TACGACGGCGGCAAGCGCGCGGTGTACGCGGTACGGACGGGACACCGGCACCTGGACGGAC	33512
Qy	991	atcagccgcgtaccttcagacgacacgcgaacgcgaactcttggttacagggcgtcaact	1050
Db	33511	ATCAGCCGCGCGGGAGCGCGACAC-----CTACTACGGGTTTCAGCGGCTTGACC	33451
Qy	1051	atcgaccgccagcacccgaaacacgataatggtggcaaccagatcgtggtggccggac	1110
Db	33460	GTGACGCGGACGCGCGGGACGGTGATGCGACCGGCTACAGTCTGTGTGTGGCGGAC	33401
Qy	1111	accataatctttggagacacgcgcgcgtgacgctggacgtgagcgcggtatggattggacg	1170
Db	33400	ACGCAGATCTTCGCGTCGACGGACAGCGCGCGGACATGGTCCACGGCTGGAGCTACACC	33341
Qy	1171	agttatcccaatcgaaagcttgcgatatgcttgacatttcggcgagaccttggctgacc	1230
Db	33340	TCGTACCCGCGGCGGAGAACCGGTACACGATGGACGTCTCTCTGTCCTCCCTGGGCTCAC	33281
Qy	1231	ttcgggtctacgcgcgaatctctccgtacccagtcgcgaagctcggctggtgatgaaagcg	1290
Db	33280	TGGGCGCGAACCCGCGCGCCCGGACGACACCCGAAAGCTGGGTGGTACGCGAGGCT	33221
Qy	1291	atggcaatcgatcgttcaactcttgatcgtgatgctctcaggaacagcgcgacgttgtac	1350
Db	33220	CTGGAGATCGACCGCTTCGATTCGGACCGGATGATGTACGGCACGGGCGGCACGGCTCTAC	33161
Qy	1351	gcaacaaatgatctcacgaagtggactccc---ggcgccagattcatatcgcgccgatg	1407
Db	33160	GGCAGGAGAACCTGACGAACCTGGGACACGAGGGCGGCACGTTCCGCGCTCGAGCCGATG	33101
Qy	1408	gtcaaaagattggagagacggcggttaacagatctcatcagccgcgctctggcgcccg	1467
Db	33100	GTGCGGGCGCTGGAGAGACGCGCGCTCAACACCTTGGCTCTCCCGCCGCTCGGCTCGCCG	33041
Qy	1468	ctcatcagcgtctctggagacctcggcgcttcaaccacgcgcacgcttactgcgctgcca	1527
Db	33040	CTGCTACGCGCCCTGGGTAGAGTCTGGCGCGCTTCGGGCACACCGACCTGACCGAGGTCCG	32981
Qy	1528	tcgacgatcttcaagtcacgggtgttcacgaccggcacagcgtcgactatgcggaattg	1587
Db	32980	TCGATGATGTACACCTCGCCCAACTTCACTTCGACGACGAGCCTTGGACTTCGCGGAGCG	32921
Qy	1588	aatccgtgatcatcgttcgcgtctggaagtttcgatccatcgagccaaccgaacgacag	1647
Db	32920	AAGCCGGACGTCGTGGTTCGGGCGGGCAATCTCGACT-----CGGTCCG	32876
Qy	1648	cacgtcgcgtcttcgacagacggcggaagacttggttccaaaggcagcgaaactctggcg	1707
Db	32875	CACATCGGCTTCACGAGAACGCGCCCACTGGTTTCGGCGGCACCGACCCGTCGCG	32816
Qy	1708	gtgacgacggcgcaacctcgcgcgcatcggccgcagcgctctgttctgtgtggctccc	1767
Db	32815	GTGAGCGCGCGGACGCTGTCGCGCGGGCGCGGACGCGCGCTGCTGTGGAGCCCG	32756


```
/protein_id="AAK78896.1"
/db_xref="GI:15023821"
/translation="MPFINSTVTVKLDKEXESTLKVELGKII IELLPGKSDMLMVGFK
DNVPLXPKGKKKAAPIEVKIFGSADKSKNLVSAISSLMENELSI PKDCIYIVD
EVSNAWNGSLF"
3104..4252
/gene="CAC0921"
CDS
3104..4252
/gene="CAC0921"
/codon_start=1
/transl_table=11
/product="Metal-dependent amidohydrolase"
/protein_id="AAK78897.1"
/db_xref="GI:15023822"
/translation="MTIEKQEEFLKLGIRRELHKHPELSMKEYETTKRIKRWLYE
NDIKVLDPLFVGVGAIEGHEGKTVAIRADIDALPIEKTLPSPSSVNGKIMHAGG
HDHPTAALITAILSKRAEITGRVRIIFQPGETGKGSQYIIIEGLVLDKIDCFGM
HNPDPVPGTIGIRSELMAVDKFEIKVFGSGHAGIPNKCIDIIVVASQIVASQMT
VVSRSISPIESVISTIRFNSGTTWVSDSAEMEGTVRTFNNIRIEIRNLHRTAK
GIARALGGIEFKWFYSYSPVLNDVRFKLCIEAAEDNGYKYVEAKRNGLGGEFSFVQ
TVVPSFVWVGVDGTEBWHKANYNLKEDALMVAANYFSDLALVLSK"
4594..4863
/gene="CAC0922"
CDS
4594..4863
/gene="CAC0922"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B. subtilis"
/protein_id="AAK78898.1"
/db_xref="GI:15023823"
/translation="MFDKAALECEDIKKSTKSTQTLIDSWYGESCTAFKTYLILDR
KLNTYTEILOEYSKALYDISETQNTDGDIAKAMSKAVNEASEQ"
5056..6348
/gene="CAC0923"
CDS
5056..6348
/gene="CAC0923"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein; CF_1 family"
/protein_id="AAK78899.1"
/db_xref="GI:15023824"
/translation="MQSLKVDNRNGLLKINEAKOGLKSAVDKFSLNKRSIFIELKN
LKRRLAVNLEGEIRKINSIDYMKRFTDVKDKASKLKTSGYTVRKVGLNIFGE
EVERVPGIGVPIIDAVESVDRNKFVKKTITITLETGAIGCVITVPEAGITWKA
WDALSAAGGVSDNLTGAASAIYQKRVKSEEDSGYDLFOSIFGGAGWSAKTILG
VKPNSWCSSGKTYLDRTSATVTVLVNHSVTSLSGDKVEKLPALKKAESSANKTLEA
ANNAEKIHDMMATKNGKINGINNAKREVVDLKLKKNVGKTNFEPYKKNFYKDKNLL
KDKGNLPQLQDOLLIQDQOQLETFKNTUNAKNKNKSIDRLNKNYRDTIKASLYNIPA
GGTYLNTEDYQGSFHFNYFPGEIKPIFEIKVVLGGER"
6345..6923
/gene="CAC0924"
CDS
6345..6923
/gene="CAC0924"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78900.1"
/db_xref="GI:15023825"
/translation="MIRPQRMAIFLTLSILLICGDIKIIISFGIEFFDILCNLFPI
GLGIIGRIYRWYKQIKRYTELLKEDLDKIEYIIYDIERVGVNTAFYIDGVKC
KLRFQYKCKDNNDYIYLEPIYKNGEKFDFIAYADKKNDYKKSISTNEITECEEK
AEKKRRKRLNEVNSRMELAKKYEAKNM"
7003..8043
/gene="CAC0925"
CDS
7003..8043
/gene="CAC0925"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78901.1"
/db_xref="GI:15023826"
/translation="MENENVLRAQELNNRGLIFIRNAGDEALKYFNKAIKEDENPKE
```

```
AYLNKAALCLAMNKVDDAVSCYDKLIIKYPNEGKAYFGKGNVFFYKDNIKKATIELYN
RSYLGMSGSEVYVNLALCMESIGELDDAIKLNRLAIINDQNIITUNKAAVLIKIK
KFEALCEYNRVLKIEVDNEEVHFKAMLLGEMGKIEKAFETISRGFALLGEOMTLCY
DKAVLEKQFEKALECTEKALLFDEGNVLLILKKGDLTYLKKIDEASTYDKILE
FRENNEGVPFAKANLCMLLGEHDAEEFFREITERUSGEESVINSIYRALNLKKA
KTRKQRRLIRKQ"
7998..8357
/gene="CAC0926"
CDS
7998..8357
/gene="CAC0926"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78902.1"
/db_xref="GI:15023827"
/translation="MGKNEAKAYEAKIKKNYFLIKYDYDKVKNLFFKANCRLDIED
YKAEDEYFIIDNDDALEVILMRANRISLNKYEAAKEDLNTVIKLNPAIRKRVIEL
DEGLKEFLSGKSECYR"
8379..8675
/gene="CAC0927"
CDS
8379..8675
/gene="CAC0927"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B. subtilis"
/protein_id="AAK78903.1"
/db_xref="GI:15023828"
/translation="MNSDEIRIDEKACDDVIKDKAVGEKVKETYNLIHLSNRHIEDG
IKGSADVDSITTEYNKIVKVKEMQDNMGNDVAANNIVSSFDAEDKMKTKFYR"
8713..9012
/gene="CAC0928"
CDS
8713..9012
/gene="CAC0928"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B. subtilis"
/protein_id="AAK78904.1"
/db_xref="GI:15023829"
/translation="MVDRCIDIEIASIFEETIKVYSDSKVRLHDILCDLENELKRWED
TWEGDAKKKEDSTFPGFYSAMKKDCDMLLELIKELKLVKTSFESLDEMNKLNKK"
9712..10335
/gene="CAC0929"
CDS
9712..10335
/gene="CAC0929"
/codon_start=1
/transl_table=11
/product="SAM-dependent methyltransferase"
/protein_id="AAK78905.1"
/db_xref="GI:15023830"
/translation="MGSMOYFDNIAASWNVIRSEYFEERLYKVKLSVKITADIF
GCCTGFTSLALASDANNVGLDNMNLRELKKSANDKLNLIYLLKSSITSVTLFDE
SIDAVFINNALHIKDAEKAIEKMYRVLKDKGTIVISDVSEHNGKKNAREMPFENLGF
SNEQIDMWLNKGVGKFNIDIENTDLCKGTSSRGEHTETSIFIATAAK"
10353..11498
/gene="CAC0930"
CDS
10353..11498
/gene="CAC0930"
/codon_start=1
/transl_table=11
/product="Cystathionine gamma-synthase"
Query Match 13.1%; Score 375.6; DB 1; Length 12732;
Best Local Similarity 50.8%; Pred. No. 2.8e-33;
Matches 1121; Conservative 0; Mismatches 1049; Indels 36; Gaps 8;
Qy 154 tacacctggagcaacgtggcgatcgggcgcgcttcttcgcgacggatcgcttcaat 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 TATAAGTGGGATAATGCAAGATAGGTGCTGGTGATATGTTCCAGCAGTTATATTAAT 249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 214 gaaggtgcacccggaatctctacgtgcgacggacatcgggggggtatcgatcgatggat 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Db 250 AAAACAGAAAAGATTAAATATATATGCTCGTACAGATATGGAGGACATATAGATGGAC 309
QY 274 gccgcgaacggcggtggatccctctctctgtgattgggtggatggaaacaaattgggggtac 333
Db 310 AAGGCTAAATAAAGTGGATTCCTCAATTTACTGAT-----GGATTTTCTGATTTGGACAATG 363
QY 334 aacgggctgtcagcattgcggcagacccgatcaataactaataaagtatatggcgccgctc 393
Db 364 CTTTGGCTGCGAAGATATAGCAACTGACCCCAATAGACACTAATCGTGTTTATATAGCAGCA 423
QY 394 ggaattacacacacagctgggaccacaacacgagcgagcgattctccgctgtctgtatcag 453
Db 424 GGTTTGTACACAAATGATTGGCAAGATGAAATGCAATATATTCTTAGTTCCACAGGATAAA 483
QY 454 ggcgcgaactggcaataacgcgccctgcgcgtctcaagcttgccggaacacatgcgccggcgct 513
Db 484 GGTAAACACATTTGGAAGATACCAACATTTCCATTTAAGGTAGAGGTATATATGCGGGGAAGA 543
QY 514 ggaatggcgagcgcttgcgttgcgttgcgaagcttctgcgaacatgacaacattctgtattcgcgcgc 573
Db 544 AATATGGGAGAAAGACTTCAATATAGATCTTAATGATGATAAAATATCTTTATCTTTGGAGCT 603
QY 574 ccgagggcaaaaggcttggagaacacagattccgggcgcgacctgggtcccgagatgacg 633
Db 604 AGAAGTGGTAATGGACTTTGGAAGGTGAAGACTATGGACAAACATGCTCTAAGGTTGAT 663
QY 634 aactttccgagcattagcagcattgcaattgcaaatcccaactgacacgacccgctctacagagc 693
Db 664 AATTTCCCTGATACAGCGACTATGTACAAAGTCC-----TCAAAATGAATATACAGCA 717
QY 694 gatattcaagcgctgcgttgcgttcttcgacaagcttctgcacatcttcgcatcgctcggaagcg 753
Db 718 GATAAAGTTGGAGTAGTTTGGGAACCTTTGATCTTCAACAGGAACAAAGGAAGCCCA 777
QY 754 agtaagaccatttttggggcttggcggat-----cccaataatccggtctctcttgagcaga 810
Db 778 ACTCAAACTATGTATGTGTGGTGGCAGAGATAAAGACAGAGAAACAATAATTTACGTTACTAAT 837
QY 811 gacggcgcgacgcttgacgagcggtgcgggtgcggcagccggcttcatccgcgacaag 870
Db 838 GATGGAGAAACATAGGATGGATGAGTTAAAGGACAGCAAAAGGATATCTTCCACATCAT 897
QY 871 ggcgtctttgaccgggttcaaccgcgtctctatatatgtccaccagcaatcagggltggtccg 930
Db 898 GGTATACTTG---CTTCCGATGATGCTTTATATATCATATATAGCAATACTTGGCGACCA 954
QY 931 tatgacgggagctccggcagctctggaattctcgtgacctccggacatcgagcagcga 990
Db 955 TATGACGGCTCAGATGGACAAGTTTGGAACTATAATACCAAAACAGGAGAATGGACAAAC 1014
QY 991 atcagcccggtaccttcgacgacagcgccaacgactacttggttacagcgccctcaact 1050
Db 1015 ATAACTCCACCTCGAGTAGTGACAC-----TAAATCTGGTTTGGAGGAATATCA 1065
QY 1051 atcgacgcagcagcaccgaaacagataatgttggaaccacagatatcgtgtgcccggac 1110
Db 1066 GTAGATGCACAAATCCAAATATATGTTGATTTGCAACTTTAAATAGATGGTGGCCGGAT 1125
QY 1111 accataatctttggagcaccgcgcggtgcgacgtggagcgcggaactctgggattggaacg 1170
Db 1126 GAAGAAATATACAGAAGTACAGATGCAGGTGAAGACCTTGAACCTATATGGGATTTGGAAT 1185
QY 1171 agttatcccaatcgaagcttgcgatattgtcttgaaatttcgcggagaccccttggctgacc 1230
Db 1186 GGATATCCAAACCGTACTTTTGGGATAACTTAGATTACTCAGCACAAACCTTGGCTTTGAT 1245
QY 1231 ttcggcgtacagcgc---aatctccgtaccagctccgaagctcggtggatggatgaa 1287
Db 1246 TGGGTTAAGACCGGGTTAACACCACAGATCCATTAGTTAAATTAGGCTGGATGGGA 1305
QY 1288 gcgatggcaatcgccttcaactctctgcgatcgtctctacggaacagcgcgacgttg 1347
Db 1306 GACTTAGAAATTTGATCCATTTACTCGGATAGAAATGTTCTACCGTACAGGTGCAACTCTT 1365

RESULT 5

ABI292929

LOCUS

ABI292929

2579 bp

mRNA

linear

PLN 08-AUG-2000

QY 1348 tacgcaacaaatgatctcacaagaagtgggactccggcgccagagatttatatataatcgcgcgcatg 1407
Db 1366 TATGGAACTGATGATTTGATAATATGGGATAAAGGTAAAGATGTAGACATCTCTGTAAAG 1425
QY 1408 gtcaaaaggattgagagacgcggtaaacgactctcatcagccgcgctgtgcgcgccccg 1467
Db 1426 GCAAAATGGAATTAAGAATGTGAGTAATGATGTTGTAGTACCAACAAAAGGCGCACAA 1485
QY 1468 ctcatcagcgctcgcgagacccgcggccttcaccacacccgacgttactgcgcgtgcca 1527
Db 1486 CTTTTAAGCGCTGTTGAGATGACTCGGATTTTATCATGATGACATAAATAAGTTCCA 1545
QY 1528 tcgacgatcttcacgctcacccggtgttccacgacggcaccagcgtcgcactatgcggaattg 1587
Db 1546 AGTAAATGATGACACACCTTAATTTAGTGTACACAACAGCATTTGATTATGACAGAAAGT 1605
QY 1588 aatccgctcgatcgtctgcgcgtggaagtctcgatccatcgagccaacccgaacgacgag 1647
Db 1606 GTTCTCAACTTTGTTGTTGCTGTTGGAAATGTTGATACTAGTAAATCAACAGATAAA 1665
QY 1648 cagctgcgcttctcgacagacgcggaagaactggttccaa---ggcagcgaacctggc 1704
Db 1666 GATTGTGGAATTTCTTACGATGGTGGTAAGAACTGGTTCTCAGCTGGAAGCAATATATCA 1725
QY 1705 ggggtgacgagcgccgacccgctgcgcgacatcgccgacggcgtctcgtttcgtcggcgt 1764
Db 1726 GGAGTTTATAAAGCAGGAAGCTAGCAGCTGGTGTGCTGATGCTAAAACTATAGTATGTCT 1785
QY 1765 ccggcgatccccggtcagcctgtgtgtacgcagtcggatttggcaactcctcctggcgtgct 1824
Db 1786 CCAGAGGAAGGTGCTAATGACGATATAGCACAGATA---ATGGTAAATAATGACACCG 1842
QY 1825 tcgcaagggttctccgcccaatgccagatccgcctcagacggcggtggaatcccaagacttc 1884
Db 1843 TGTCTGCTTCTCCACAGGGAAGAAAGTTAGATCAGACGCTGTAATTCAAAAGAAATTC 1902
QY 1885 tatgcctataccaaatgaaccttctcgaagcagcgcgcgcgatcgacatcccaaccg 1944
Db 1903 TATGTTTCTTAATGGTAATTTCTATATTAGTACTGATGAGTGCAGGTGCAACATTTACTCAA 1962
QY 1945 gtgcggcggtcttccgacgacggtgcgcgtggtgtcatgttccacgcggtgcctgga 2004
Db 1963 AGTTCACAAACAGCTCTTCTCACAAGGCAAAAGGAATCTTTAAGACTGTTATAGTTCAT 2022
QY 2005 aaagaaggcgatctgtgctcgtgcatcgcagcggttaccacactcaacaaatggcgccg 2064
Db 2023 GAAGGTGATATTTGGATTGCTGGAGGAAAAGACGGATTATGGCATTCACATGATCTCTGT 2082
QY 2065 agcagttggtcgaatcacccgcgtatcctcgcggtgaaactgggatttggtaagtct 2124
Db 2083 GCAACATTCACAAAAGTATCTGTTGTGATGCTTCTGATACAGTAGTCTTGGAAAAATCC 2142
QY 2125 gcgcggcggtcgtcataccacgcgtcttgcgtgcgcgacgctcggagcgttacgcggg 2184
Db 2143 AAACAGATGATGGATATCCAGGTATATACATGGATGCTACAATTTGGAACAGCGGGA 2202
QY 2185 gcttaccgctccgacgactgtgggacgacccgtggtactgatcaatgatgaccagcaca 2244
Db 2203 ATATTCGTTTCAGATGATGAAGGTGCAACTTGGTAAGAATAAATGATGATGCTCACCAA 2262
QY 2245 tacggaatctggggaagaagaatcacccggtgaccacgcgaatttaccggcggtgtacata 2304
Db 2263 TATGGTTCCCGACACTACTGCATAACTTGGAGATCCTTAACAAATATGGACAGATTTTGT 2322
QY 2305 ggcacaacgcgctggaattgtatcggggacattggtggtgcgc 2350
Db 2323 GGAACTAATGGAAGAGTATTGTTTATGGAGACATAGATGTTCTC 2368

[illegible]

[illegible]


```
misc_feature 2033..2884
/gene="mana"
/notes="Region: beta-1,4-mannanase catalytic domain"
/evidence="experimental"
misc_feature 2885..2959
/gene="mana"
/notes="Region: proline-threonine linker"
/evidence="not_experimental"
misc_feature 2960..3409
/gene="mana"
/notes="Region: cellulose-binding domain type IIIB"
/evidence="not_experimental"
CDS 3894..4400
/notes="orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF22275.1"
/db_xref="GI:6651328"
/translation="MGASGVFSTAGNGYGLKARCKRRCVGRNLLGHLRFVVDDEF
RFAGARKRLGRFSPNRNLGEAQNFSGTFTVOPDDRRIRRRNGOLFVRFQASI
GPKRLTEIAEQARAKQASDPQLKNVILHFDIHPYGVRRMDPFPHTEKOKS
RSILTRY"
BASE COUNT 1040 a 1276 c 1439 g 812 t
ORIGIN
* Query Match 3.4%; Score 96.4; DB 1; Length 4567;
* Best Local Similarity 57.1%; Pred. No. 0.027;
* Matches 216; Conservative 0; Mismatches 156; Indels 6; Gaps 2;
QY 2494 ccgagtcgtctcggttcggtccaccatcggtcgccgagccggtcttcgtccaccagc 2553
Db 2837 CCGACGCGTGGGGGACATGGTTCCGGAGCAACCGCATCGGACCTCGACTTCGCCGACA 2896
QY 2554 ccgtcttcgtccacgtcttcgtcgccgagcccaacgcgctcgctcgccggtgctcggt 2613
Db 2897 CCGACACCTACACCGCGCGCGCAACCGCGACCGCGACCGCGCGTCCGCCGCG 2956
QY 2614 gggg---tgaagtgagatagaataatgattcgcgccggtgataatcgatcaag 2670
Db 2957 GGCAACTTGTCTACAGTACCGCGCGCGGACACGCAACCGCGGACCAACCGCTGAAG 3016
QY 2671 ccgggttcgagtggtggaataccggtcgctcggtgattgctgcagcgtgacggtg 2730
Db 3017 CCGCATTTTCGGATTGTAACCGCGGACGTCGAGCGTGGCGTGTGCGGACTGACGATT 3076
QY 2731 cggtaactggttcacccggggtggtggtcgctgcacactggtgtacaaactgtgactggcg 2790
Db 3077 CGGTACTGGTACACGGGTGGACGGGACAAACCGG--CAGGTGTCAACTCGGACTGGGCG 3133
QY 2791 gcgacggtggtggaatatccgcgctcgttcggtcggtgaacccggcgagccgacg 2850
Db 3134 CAGGTGGGTTGCTCGAACTTCGCGGGGACGCTTCTGTAAGCTTTTCGACGGGCGGACGGG 3193
QY 2851 gcgacacccactacgtcga 2868
Db 3194 CGGACTACTACATCGAG 3211

RESULT 11
AL645600/c
LOCUS AL645600 183665 bp DNA linear HTG 01-FEB-2002
DEFINITION Mus musculus chromosome 11 clone RP23-396N4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL645600
VERSION AL645600.8 GI:18491381
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (sites) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
```

```
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 4, 2002 this sequence version replaced gi:18476868.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm396N4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 183201 bases at least Q40
Consensus quality: 183287 bases at least Q30
Consensus quality: 183384 bases at least Q20
Insert size: 183465; sum-of-contigs
Insert size: 176789; 3.5% error; agarose-fp
Quality coverage: 13.90x in Q20 bases; sum-of-contigs Quality
coverage: 15.10x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source            1..183665
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="11"
                        /clone="RP23-396N4"
                        /clone_lib="RPC1-23"
     misc_feature       1..3150
                        /note="assembly_fragment:05886"
     misc_feature       3251..60564
                        /note="assembly_fragment:00037"
                        /fragment_chain:1"
                        60665..183665
     misc_feature       /note="assembly_fragment:06038"
                        /fragment_chain:1
                        clone_end:SP6
                        vector_side:right"
BASE COUNT 51045 a 35803 c 35999 g 60618 t 200 others
ORIGIN

Query Match 3.2%; Score 91.4; DB 2; Length 183665;
Best Local Similarity 56.0%; Pred. No. 0.04;
Matches 173; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 2347 gcgccgtccggatccggtctccggtcggtgagtcgctcgctcgccgagcctgagccg 2406
Db 127835 GCTCCGGCTCCAGTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127776
QY 2407 agcccgagcccgagcagctcgccatcgccgtcgccgagcgtcgagtcacatccctcg 2466
Db 127775 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127716
QY 2467 tcgccgtctccgtcgccgtccaccatcccgagtcgctctcggtctccgtccaccatcgccg 2526
Db 127715 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127656
QY 2527 tcgccgagcccgcttcgtcacccgagcccgcttcgtcacccgcttcgtcgccgagccca 2586
Db 127655 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127596
QY 2587 acgccgtcgctcgccggtggtcggggtgaggtgcagatataagaaatgattcg 2646
Db 127595 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127536
```


Oy 2647 gcgcgggt 2655
 || |||| |
 Db 127535 GCTCCGGCT 127527

RESULT 12

AP003605

LOCUS

DEFINITION Nostoc sp. PCC 7120 plasmid pCC7120epsilon DNA, circular BCT 05-DEC-2001

ACCESSION AP003605

VERSION AP003605.1 GI:17134931

KEYWORDS

SOURCE

ORGANISM

Nostoc sp. PCC 7120 plasmid:pCC7120epsilon DNA.

REFERENCE

1 (sites) Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

AUTHORS

Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
 Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
 Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
 Nakazaki,N., Shimo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
 Yasuda,M. and Tabata,S.

Complete genomic sequence of the filamentous nitrogen-fixing

TITLE

Cyanobacterium Anabaena sp. strain PCC 7120

JOURNAL

DNA Res. 8 (5), 205-213 (2001)

MEDLINE

21595285

PUBMED

11759840

REFERENCE

2 (bases 1 to 40340)

AUTHORS

Kaneko,T.

TITLE

Direct Submission

JOURNAL

Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research

Institute, The First Laboratory for Plant Gene Research; Yana

1532-3, Kisarazu, Chiba 292-0812, Japan

(E-mail: kaneko@kazusa.or.jp,

URL: http://www.kazusa.or.jp/cyanobase/.

Tel: 81-438-52-3935(ex.2338), Fax: 81-438-52-3934)

Location/Qualifiers

1. 40340

/organism="Nostoc sp. PCC 7120"

/plasmid="pCC7120epsilon"

/db_xref="taxon:103690"

/note="synonym: Anabaena sp. PCC7120"

1429. 3222

/gene="alr9001"

/note="alr9001"

/note="ORF_ID: alr9001

similar to mobilization protein"

/codon_start=1

/transl_table=11

/protein_id="BAB77487.1"

/db_xref="GI:17134932"

/translation="MVAAIILRVKLSFGNVGSGKHTARLQDTPNADTTKKNIRLI

GMEDNSLEDVYKTKTANTKKPKDAVLCSPMFLSASPEYFRDDPSRAGWNHER

MVLFACASRWLINNGDKCYRAELHDEATPHIAYIYPINDTKQLSKHEMFGNG

RVGSIKSLQDSIALAPLGERGVGSKATHTVKEYIQAVNSEPLTAVSNKLL

APQLESATNYVARIQNDQFHAINQLADRAPMQERLERAQRASEKQRLEKE

VRSELEKTLQRLDELVDVWELGDIYERWRGHGHIINIDGSKFYDFSPQKGGG

GAIDLVMHYNQCNFRQAVVWLHERFGEAGVERAAIAHVKNRAADIQAEPQPTPV

ENKANMPAVERVYLTQQRIPSDCLQMLGLLYADQQNAVFMRLDLEGTSQAFLR

GNPCENSPKGYSKGPKRSDSMFYGLGGQATDKVAHVLLCSPIEAVSRAMLEYVR

GNVPPRTLYMVDNINSLPVERLQVNPVILVTFGDKQSTNAQAQVLELLPQSQQVL

SKAADNYYQLLEYGWLRRQQYQQQDDLSI"

complement(3399..3833)

/gene="alr9002"

complement(3399..3833)

/gene="alr9002"

/note="ORF_ID: alr9002

unknown protein"

/codon_start=1

/transl_table=11

/protein_id="BAB77488.1"

/db_xref="GI:17134933"

gene

CDS

/translation="MRKLVPSTERYRELKTAQAQEEVEVITDVERYSDLPOLKTI
 IFELHNKSGMVNVATGAFIPYSFAATTINYSALYIFGGAGAGATFGVIVGGPVGAA
 VGGIGAIIVGTAVATLNGKHVDIEINAGKLRFRKISPSIK"
 complement(4726..8019)
 /gene="alr9003"
 complement(4726..8019)
 /gene="alr9003"
 /note="ORF_ID: alr9003
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77489.1"
 /db_xref="GI:17134934"

/translation="MVAQSSSTAIAGSDSFDIRNFIDQLEPTRVKNKYICPVCGGHN
 LSINPNNGKSYCNECLHRDIREAIPWTQVLEERKLGSTLSPKPLPIKAKKPATVPK
 VLDVPSQLRIICLLGEITTPQVTPDFPKSVAIRLSDSGATSQLKETKEYDYDG
 NGRKAHRFSCPCAAAPKGRVKTFSVSDIPTNKNVAKKGGFWPAVQSEAIATIKAT
 DGPVLLAHEGKCVASRLNLSATITWVGSSSDRLILSLTQIQTSTKDFLLATCV
 DNDSITWNKQRIKEICQAGVSFAIDLLKIQPNLDQKGDIAIDIESGTMGDELVEL
 LUKQIEITQKQQSAQDTLDVADLDEEDPSDFLQKGFNTLYGETKWTACDGLVY
 WHGKYRSPDAVERRRITNYCNSGPIPKNGDSYVITYPAKPNKVKLELWVKLRY
 EIDPQLNPNVGNCTNGVLAIDWSGPIPRPVEHPDKHFTYEPKLVKYPKADSSH
 CDRLLCLDSPOQOILLNLGASLDIAEVRKRGPKVLLACGLGANGKDSIQVVS
 TIYHOGMTSCSLADFAVDEGRKFALALVNSRVNASENPOTTRLDKIQSLKLEAT
 GNVLSERKGDHIEFNSRAIGLFNLDTPSWTGTQAILDRIALIFRKTFTNDDP
 NNHDELLADPRFAYDVFQECVAPAFNLNKNMIAGLQALIAEGIDYISCTQQAIEIOAE
 NSHLFQCKDTGLGYKANGIVTAFDIQRLNMYIDNGTLSPFESSNGKVAIKWQS
 KPSDKTVAINOALPRIKALFPKTLVTPHPSGKRLQALQIGISFDNVPMSITGPI
 ATSENSTPNSTPNHQSLINQDFTHTTSTDEGLDEKNSGFOKVLDDQPKSEIINV
 VDRADQDHGNLESKLIDINPNDVFENEIEGDKLWVEVCGEVPASTADSTGVDIAPG
 ELTAMPESLITTEVESQLKSDVAEVLQGEAEASHQTVNVRDEQPKSASFQSPCS
 TSGSTLESLFEQKLMGNVALIRECIADESGLIAELIEETWTSEKSAVWKELTPEERQ
 AVKQLKALE"

gene

CDS

complement(8400..8570)
 /gene="parg"
 complement(8400..8570)
 /note="ORF_ID: asl9004
 parg gene product"
 /codon_start=1
 /transl_table=11
 /product="plasmid partition protein"
 /protein_id="BAB77490.1"
 /db_xref="GI:17134935"

/translation="MTDSNKQVFRVFRVDEKRDREKFTICIQKTTTMDVYLKGLLDKW
 LEENDPESTKNK"

gene

CDS

8828..9880
 /gene="alr9005"
 8828..9880
 /gene="alr9005"
 /note="ORF_ID: alr9005
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77491.1"
 /db_xref="GI:17134936"

/translation="MTDLAEMPENALPRHQGTULTIAGYDLGNSGVKFTVSDRKIRFP
 SYLENYRPTLPTGEYVEYLEGDAITKLDYKQWLSGYAAVDANPKNHLRVTTDDATA
 KYTQSLKLLAALSNYPYKPVINLIICASHERGDLDEQLIDAIGAKHIVKFGGKPIP
 TTVNHLVYDDEGHAAIAANAHTLDTSKQNVTDIGNRTVLTATLIGOKGHLANKRTE
 DNQVELIEMISVNPFTKRLYGEIAPHILRQGLSESEKPFYHQQSFEDYVRQEL
 MPWQKSLAPVFKFIHPWKINADACLIIIGGSQLPSVDEALKAKGFVIAENPLWAAE
 GLYQLATMYSRGIDE"

gene

CDS

9921..10088
 /gene="asr9006"
 9921..10088
 /gene="asr9006"
 /note="ORF_ID: asr9006
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77492.1"
 /db_xref="GI:17134937"

JOURNAL

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625700.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCSW
Center clone name: CH230-8G23
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList

Consensus quality: 148905 bases at least Q40
Consensus quality: 156308 bases at least Q30
Consensus quality: 160895 bases at least Q20
Estimated insert size: 150335; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1
10936: contig of 10936 bp in length
10937 11036: gap of unknown length
11037 17011: contig of 5975 bp in length
17012 17111: gap of unknown length
17112 21819: contig of 4708 bp in length
21820 21919: gap of unknown length
21920 27267: contig of 5348 bp in length
27268 27367: gap of unknown length
27368 32998: contig of 5631 bp in length
32999 33098: gap of unknown length
33099 38238: contig of 5140 bp in length
38239 38338: gap of unknown length
38339 42861: contig of 4523 bp in length
42862 42961: gap of unknown length
42962 48627: contig of 5666 bp in length
48628 48727: gap of unknown length
48728 52551: contig of 3824 bp in length
52552 52651: gap of unknown length
52652 56759: contig of 4108 bp in length
56760 56859: gap of unknown length
56860 60122: contig of 3263 bp in length
60123 60222: gap of unknown length
60223 64019: contig of 3797 bp in length
64020 64119: gap of unknown length
64120 67830: contig of 3711 bp in length
67831 67930: gap of unknown length
67931 70511: contig of 2581 bp in length
70512 70611: gap of unknown length
70612 74478: contig of 3867 bp in length
74479 74578: gap of unknown length
74579 77220: contig of 2642 bp in length
77221 77320: gap of unknown length
77321 80928: contig of 3608 bp in length
80929 81028: gap of unknown length
81029 83523: contig of 2495 bp in length
83524 83623: gap of unknown length
83624 86404: contig of 2781 bp in length
86405 86504: gap of unknown length
86505 89767: contig of 3263 bp in length
89768 92456: contig of 2589 bp in length

92457 92556: gap of unknown length
92557 94688: contig of 2132 bp in length
94689 94789: gap of unknown length
94790 97736: contig of 2948 bp in length
97737 97836: gap of unknown length
97837 100283: contig of 2447 bp in length
100284 100384: gap of unknown length
100385 103335: contig of 2952 bp in length
103336 103435: gap of unknown length
103436 106003: contig of 2568 bp in length
106004 106103: gap of unknown length
106104 108849: contig of 2746 bp in length
108850 108949: gap of unknown length
108950 112359: contig of 3410 bp in length
112360 112459: gap of unknown length
112460 116805: contig of 4345 bp in length
116806 116905: gap of unknown length
116906 118533: contig of 1948 bp in length
118534 118953: gap of unknown length
118954 121180: contig of 2227 bp in length
121181 121280: gap of unknown length
121281 123672: contig of 2392 bp in length
123673 123772: gap of unknown length
123773 125124: contig of 1352 bp in length
125125 125225: gap of unknown length
125226 126338: contig of 1114 bp in length
126339 126438: gap of unknown length
126439 127933: contig of 1495 bp in length
127934 128033: gap of unknown length
128034 129916: contig of 1883 bp in length
129917 130016: gap of unknown length
130017 132151: contig of 2135 bp in length
132152 132251: gap of unknown length
132252 133407: contig of 1156 bp in length
133408 133507: gap of unknown length
133508 134852: contig of 1345 bp in length
134853 134952: gap of unknown length
134953 137209: contig of 2257 bp in length
137210 137309: gap of unknown length
137310 138426: contig of 1117 bp in length
138427 138526: gap of unknown length
138527 139723: contig of 1197 bp in length
139724 139823: gap of unknown length
139824 141817: contig of 1994 bp in length
141818 141917: gap of unknown length
141918 143387: contig of 1470 bp in length
143388 143487: gap of unknown length
143488 145552: contig of 2065 bp in length
145553 145652: gap of unknown length
145654 147296: contig of 1644 bp in length
147297 147396: gap of unknown length
147397 148991: contig of 1595 bp in length
148992 149091: gap of unknown length
149092 150398: contig of 1307 bp in length
150399 150498: gap of unknown length
150499 152456: contig of 1958 bp in length
152457 152556: gap of unknown length
152557 154951: contig of 2395 bp in length
154952 155051: gap of unknown length
155052 156913: contig of 1862 bp in length
156914 157013: gap of unknown length
157014 159194: contig of 2181 bp in length
159195 159294: gap of unknown length
159295 161394: contig of 2100 bp in length
161395 161494: gap of unknown length
161495 163410: contig of 1916 bp in length
163411 163510: gap of unknown length
163511 164613: contig of 1103 bp in length
164614 164713: gap of unknown length
164714 166031: contig of 1318 bp in length
166032 166131: gap of unknown length

Query Match

3.0%; Score 86.4; DB 2; Length 183544;

[illegible]

RESULT	14
AC099417	
LOCUS	
DEFINITION	AC099417 linear DNA 97225 bp HTG 20-DEC-2001
ACCESSION	Rattus norvegicus clone CH230-2M18, *** SEQUENCING IN PROGRESS ***,
VERSION	51 unordered pieces.
KEYWORDS	AC099417
SOURCE	AC099417.2 GI:17940849
ORGANISM	HTG; HTGS_PHASE1. Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE AUTHORS	TITLE JOURNAL REFERENCE
1 (bases 1 to 97225)	1 (bases 1 to 97225)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denna, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhardt, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Lousaged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Oanh, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokhan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Slisone, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villaion, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.	2 (bases 1 to 97225)
2 (bases 1 to 97225)	2 (bases 1 to 97225)
Worley, K.C.	Worley, K.C.

Direct Submission
 Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:16930824.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GABK
 Center clone name: CH230-2M18
 ----- Summary Statistics -----
 Assembly program: Phrap; version 0.990329 First call to
 findPhrapList
 Consensus quality: 74561 bases at least Q40
 Consensus quality: 81622 bases at least Q30
 Consensus quality: 88369 bases at least Q20
 Estimated insert size: 71621; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 51 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 5462: contig of 5462 bp in length
 5463 5562: gap of unknown length
 5563 1113: contig of 5569 bp in length
 1132 1133: gap of unknown length
 1133 14471: contig of 3240 bp in length
 14471 14571: gap of unknown length
 14571 18545: contig of 3974 bp in length
 18545 18645: gap of unknown length
 18646 19789: contig of 1144 bp in length
 19790 19889: gap of unknown length
 19890 21984: contig of 2095 bp in length
 21985 22084: gap of unknown length
 22085 24323: contig of 2239 bp in length
 24324 24423: gap of unknown length
 24424 25964: contig of 1541 bp in length
 25965 26064: gap of unknown length
 26065 28007: contig of 1943 bp in length
 28008 28107: gap of unknown length
 28108 30361: contig of 2254 bp in length
 30362 30461: gap of unknown length
 30462 32510: contig of 2049 bp in length
 32511 32611: gap of unknown length
 32611 34213: contig of 1603 bp in length
 34214 34313: gap of unknown length
 34314 37129: contig of 2816 bp in length
 37130 37229: gap of unknown length
 37230 38418: contig of 1189 bp in length
 38419 38518: gap of unknown length
 38519 39956: contig of 1438 bp in length
 39957 40056: gap of unknown length
 40057 41230: contig of 1174 bp in length
 41231 41330: gap of unknown length
 41331 43240: contig of 1910 bp in length
 43241 43340: gap of unknown length
 43341 45179: contig of 1839 bp in length
 45180 45279: gap of unknown length
 45280 47144: contig of 1865 bp in length
 47145 47244: gap of unknown length
 47245 49086: contig of 1842 bp in length
 49087 49186: gap of unknown length


```
gene
complement(2104. .2715)
/gene="lef-2"
CDS
complement(2104. .2715)
/gene="lef-2"
/note="ORF4"
/codon_start=1
/product="LEF-2"
/protein_id="AAK85568.1"
/db_xref="GI:15213129"
/translation="MDNVAWNPSSANTATIKKNETYLDLPNDPMDVLTLSPTVYVPRRG
OLVEISGMRLFALTMSKTEPKQFARRSRNICLKACADGSQLHKSLSLGLKPLPLC
IAKIMAEISNASVPRGMYKRFEPCNVLGNVLTCTCKCSACLIGALLHFFYKMDAKCV
GEVTHLLKSDOVKPSNCAMKSMKLCQPANMCKGLNPGN"
CDS
complement(217. .3070)
/note="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAK85569.1"
/db_xref="GI:15213130"
/translation="MIYRYIYFFVFRESRTKSKKPRISHIKLFLFYATLHYNNRR
INAPVMVAGHDYDREQLKRDINSRHSVHELKRSSTTGFCNRLLESSTDTAKPTVV
IKTTAAEQHTCDKL"
CDS
2990. .3424
/note="ORF6"
/codon_start=1
/product="unknown"
/protein_id="AAK85570.1"
/db_xref="GI:15213131"
/translation="MADARFLRSTRLTEYKENVVAQIDHLTNLRLINAKTTIANV
RFRFVGNAFVACMDIVHTYASSDDKITLQEPOLYFRVCRFCNEIADVPDDHS
ITRYLCAACGCLVDNPLNVETEDGVQRFMEIQRINAGGE"
gene
complement(3426. .3935)
/gene="ptp-1"
CDS
complement(3426. .3935)
/gene="ptp-1"
/note="ORF7"
/codon_start=1
/product="PTP-1"
/protein_id="AAK85571.1"
/db_xref="GI:15213132"
/translation="MFPDRWHDYTVCGNVIEGTNFCFVPLREEMFEYVTDNDLRWT
VKHLKQCCSLGAVIDLNTFRYDANMRDEGLLYKKIRVPGVLPDENVQLEHTVDVFN
VKNFTARGCPGLIGVHCTHGLNRTGYLCVRYMINILINISPDQAIARFETARGHKTERQ
NYIQHLHL"
CDS
4168. .5145
/note="ORF8"
/codon_start=1
/product="unknown"
/protein_id="AAK85572.1"
/db_xref="GI:15213133"
/translation="MSLSKLLVYDYGNMCGVNEQYGAHYLKYIVDEHITNSYVG
DYPWMHREIDTARQILSGALCVNDACHMLDLDTKRLATYKCGNTKGVCHEVYVVL
KAIDLVPIDKRSGQGVDDVDFDTILLNTKLPVLHGYLRFVYQKLEHVTDFVN
PSDRTVGMWYKFCVLTMYRIINKSPAEALLTLQKAVVEYIQPNAGDKYCAEVMA
NYVGRFCGIGKHQFAEHTANMHIMFKSMRGOTTWMDERYKSFVSKDFSRCKETYL
DLKSHVDSIYINCTDKOKNAMEFDLLCCQSEQDLDDVCDYIVNMLFFK"
CDS
complement(5163. .5516)
/note="ORF9; unique to Eppomnpy"
/codon_start=1
/product="unknown"
/protein_id="AAK85573.1"
/db_xref="GI:15213134"
/translation="MSSNSFMVDVNTHENVTNVKDTLLKPINWWSNDPSSQARCEK
GAPPYFAYKCVPLESPURGNQYKIDPTLPALOKITTTDTANKTPVLASVKPNNYQDN
TVAEAYNNYKNNII"
CDS
complement(5531. .6490)
/note="ORF10"
/codon_start=1
/product="unknown"
/protein_id="AAK85574.1"
/db_xref="GI:15213135"
/translation="MFDMMFGWFSNEQHQHHPLANHFNAODYKOYAVDRRAQSDLV
NRSVPKCHPFTFKERYIDDDNGRCRRVDFCKGLEINHMLNCKWDSKHYHLNEI
```

Query Match . 2.9%; Score 83.6; DB 14; Length 118584;
Best Local Similarity 57.1%; Pred. No. 0.32;

```
gene
complement(6427. .7182)
CDS
complement(6427. .7182)
/gene="lef-1"
/note="ORF11"
/product="LEF-1"
/protein_id="AAK85575.1"
/db_xref="GI:15213136"
/translation="MATWYTLDRVNLWNSYNNNSRKAFTMLKQRIWHYDRHFDN
ALQLLNFINSNSVDVHVKPLDDEGREWVIDADFKNYVNTKTEMLKINVGATALLLFF
TEMHVHRIFMFSGNRGFMWLKPADGFKALKSREHMYRIFEKTRLNMRDIRGSE
AHCQRAVMYMQDQILQSSKPDLELILHFWDKDFVFCNPTKQIRAPFSYNHKGTOFS
RCITQOLLEKIQKSCGSGDASTSISNITH"
gene
7254. .8732
/gene="egt"
CDS
7254. .8732
/gene="egt"
/note="ORF12"
/codon_start=1
/product="EGT"
/protein_id="AAK85576.1"
/db_xref="GI:15213137"
/translation="MHFAVLLTLLAAQAAHANILAVLTPPAYSHHLVYKRRVQVAAANK
CHNVTVIRPQLLNVYDSNAHMGCHIEQLEADMSBQYKRLVSSGAFKAGYVAADETT
TVANNYLMDFDKQNNVNVQQLNRTDFVVVEAFADYLVFGLFRPAPVIO
IAPGGLDENFDTAGAVRHPKYYPNIMRSTFVGCTAGALSEWRLYNFELLARYSDK
LLKIQFGPTPTPIRELRNNVQLLLNLHPYDNNRPVPSVOYLGGLHLTQTQPNL
QKTLERQLNASVNGIYVSGSIDTKSIHDEFQLQMLVNTFTGLNNRTVLMKVVDAVY
DSIKLPNVITONFNRQAVLHKNTVAFVTQGGLOQSDDEALHAQVPMVCLPMMGDQF
HANKLEFFKVRTLTVDVSASQLIAITDVIVNKKIYQTRMAELRAVIDYDEIAPA
DKAIKTERVIRKFGHDITYPARSLKSPAANLDQSNVFSFPL"
CDS
8914. .9528
/note="ORF13"
/codon_start=1
/product="unknown"
/protein_id="AAK85577.1"
/db_xref="GI:15213138"
/translation="METPKMLPSSFATPKRAALGTFVKTVTVTVTTSVSGRLTRDERNR
TFQVLAQLKRTLSETKLTQLQKVRNMQRVLRKKNIIADLAQLERRCRSGSK
YFAVICQNGVLITISGQGFVRQRVANMCAVGEQIFCERRNDCHADROLIAEALAA
LGSVDTISAANKRFKILLNVKVVNAKFIVQOOTLHNGFNPNYPTH"
CDS
9497. .10123
/note="ORF14"
/codon_start=1
/product="unknown"
/protein_id="AAK85578.1"
/db_xref="GI:15213139"
/translation="MDLTITLIPISLKNIIEPKRSECFKLTSMREDAEFCFLNVKCRSP
FAGKVLISITFNNEHLQATVCSQVDSVIMNPSQOEIIFDGTIKHDEGVTPVFI
VGPLFSLCESVNMHVRAVDAIRLQVTKVFINVYAEYLGKARTLKLKLFSDNNESD
LVNNIVNFNDVKIECEKYVNTKVPAINVYTGKQLTLVLFIFKN"
CDS
complement(10162. .11235)
/note="ORF15"
/codon_start=1
/product="unknown"
/protein_id="AAK85579.1"
/db_xref="GI:15213140"
/translation="MDCVANQVCSGTLPIVYVNTRDLEDCLRNKITAKIGLRFFKDFCEA
VTADKSGFLVLSGAAACHIDEKDRMLKCLDFDYNCSEKWLQAOQMKHLOACVDV
AFDNLQTLTANIRMSPLTILKCFQNGAYCFGDNDDVNLRLPLHIDTIRTFNHFNEFDLI
REFALQVEMWYQNGVDEYNKMLINKRHAVIENFVFNVRMKHPLFFGRCRCKKIFMF
GDNYHVLVSLPRLVNDQIMCLLKDIFTDLKLEFKVERRTLIRALFAKLPODACDVCV
NNHSDVAVGKHKRFENITSECKTLDIYGALGCRKLIVAYLTTNTFLNIPHYVADYA
NYPHNRQSCQEQKREFMGVNYLL"
CDS
11242. .11562
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:38 ; Search time 307.67 Seconds
(without alignments)
16010.100 Million cell updates/sec

Title: US-09-917-376-2

Perfect score: 2869

Sequence: 1 atgagatgttcgagaacat.....ggcgagacacctacttcgagn 2869

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	100.8	3.5	390	13	AAQ21833
C 2	100.8	3.5	390	14	AAQ36859
C 3	100.8	3.5	390	22	AAF76910
C 4	89	3.1	784	21	AAQ47999
C 5	89	3.1	2011	21	AAQ52204
C 6	86.4	3.0	37286	23	AAQ55922
C 7	85.4	3.0	777	21	AAQ44444
C 8	85	3.0	796	21	AAQ45005
C 9	80	2.8	11706	20	AAQ55661

10	80	2.8	11707	24	AAQ26525	Active cellulase p
C 11	79	2.8	4403765	22	AAI99683	Mycobacterium tube
C 12	78.2	2.7	58857	21	AAA58471	Nucleotide sequenc
C 13	76.8	2.7	651	23	AAQ84956	DNA encoding novel
C 14	72.8	2.5	6510	22	AAQ46648	Tumour suppressor
C 15	72.8	2.5	6510	24	ABL32395	Human immune syste
C 16	72.8	2.5	6510	24	AAQ61115	Human gene regulat
C 17	72.6	2.5	38186	20	AAQ32028	Human METHI relat
C 18	72.6	2.5	38186	22	AAQ90085	Myobacterium spec
C 19	71.4	2.5	1164	20	AAQ34250	AC004449 cDNA clon
C 20	71.4	2.5	1166	20	AAQ34248	Myobacterium spec
C 21	71.4	2.5	1166	20	AAQ34248	Myobacterium spec
C 22	71.2	2.5	494	19	AAQ64542	M. tuberculosis im
C 23	71.2	2.5	494	19	AAQ44433	Myobacterium tube
C 24	71.2	2.5	494	20	AAQ19343	M. tuberculosis an
C 25	71.2	2.5	494	20	AAQ19131	M. tuberculosis re
C 26	70.2	2.4	1596	17	AAQ22316	Nocardiosis sp. p
C 27	70.2	2.4	4403765	22	AAI99683	Myobacterium tube
C 28	70.2	2.4	4411529	22	AAI99682	Myobacterium tube
C 29	70	2.4	390	13	AAQ21833	Randomising oligon
C 30	70	2.4	390	14	AAQ36859	PCR primer for 5'
C 31	70	2.4	390	22	AAQ76910	Sequence containin
C 32	68.8	2.4	18596	22	AAQ31109	Thymidylate synth
C 33	68.8	2.4	18596	22	AAQ91215	Human thymidylate
C 34	68.2	2.4	3543	23	ABL16717	Drosophila melanog
C 35	68.2	2.4	13837	23	ABL16716	Drosophila melanog
C 36	67.8	2.4	407	19	AAQ64539	M. tuberculosis im
C 37	67.8	2.4	407	19	AAQ44430	Myobacterium tube
C 38	67.8	2.4	407	20	AAQ19340	M. tuberculosis an
C 39	67.8	2.4	407	20	AAQ19128	M. tuberculosis re
C 40	66.6	2.3	1766	19	AAQ04490	Corn p-hydroxyphen
C 41	66.6	2.3	2293	19	AAQ09659	A. cellulolyticus
C 42	66.4	2.3	6415	20	AAQ55662	DNA sequence encod
C 43	66.4	2.3	6416	24	AAQ26526	Active cellulase p
C 44	66.2	2.3	1294	20	AAQ206824	Streptomyces albid
C 45	66.2	2.3	1294	22	AAQ74537	Nucleotide sequenc

ALIGNMENTS

RESULT 1	AAQ21833/c
ID	AAQ21833 standard; DNA; 390 BP.
XX	AAQ21833;
AC	AAQ21833;
XX	08-JUN-1992 (first entry)
DT	Randomising oligonucleotide used in SPRT mRNA prepn.
DE	Systematic polypeptide evolution by reverse translation; SPRT;
XX	Ligand binding; ss.
KW	Synthetic.
OS	WO9202536-A.
XX	20-FEB-1992.
PN	01-AUG-1991; 91WO-US05463.
XX	02-AUG-1990; 90US-0561968.
PR	(COLS) UNIV OF COLORADO.
PA	Gold L, Tuerk C;
PI	WPI; 1992-080018/10.
XX	New method of systematic polypeptide evolution by reverse
DR	translation - by linking each polypeptide in sample mixt. to
PT	individualised mRNA allowing further synthesis of selected

KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.		
OS EP1033405-A2.		
PN XX		
PD XX		
XX 06-SEP-2000.		
PF 25-FEB-2000; 2000EP-0301439.		
XX 25-FEB-1999;	99US-0121825.	99US-0140695.
PR 05-MAR-1999;	99US-0123180.	99US-0140823.
PR 09-MAR-1999;	99US-0123548.	99US-0140991.
PR 23-MAR-1999;	99US-0125788.	99US-0141287.
PR 25-MAR-1999;	99US-0126264.	99US-0141842.
PR 29-MAR-1999;	99US-0126785.	99US-0142154.
PR 01-APR-1999;	99US-0127462.	99US-0142055.
PR 06-APR-1999;	99US-0128234.	99US-0142390.
PR 08-APR-1999;	99US-0128714.	99US-0142803.
PR 16-APR-1999;	99US-0129845.	99US-0142920.
PR 19-APR-1999;	99US-0130077.	99US-0142977.
PR 21-APR-1999;	99US-0130449.	99US-0143542.
PR 23-APR-1999;	99US-0130510.	99US-0143624.
PR 23-APR-1999;	99US-0130891.	99US-0144005.
PR 28-APR-1999;	99US-0131449.	99US-0144085.
PR 30-APR-1999;	99US-0132048.	99US-0144086.
PR 30-APR-1999;	99US-0132407.	99US-0144325.
PR 04-MAY-1999;	99US-0132484.	99US-0144331.
PR 05-MAY-1999;	99US-0132485.	99US-0144332.
PR 06-MAY-1999;	99US-0132486.	99US-0144333.
PR 06-MAY-1999;	99US-0132487.	99US-0144334.
PR 07-MAY-1999;	99US-0132863.	99US-0144352.
PR 11-MAY-1999;	99US-0134256.	99US-0144884.
PR 14-MAY-1999;	99US-0134218.	99US-0145086.
PR 14-MAY-1999;	99US-0134219.	99US-0145088.
PR 14-MAY-1999;	99US-0134321.	99US-0145085.
PR 14-MAY-1999;	99US-0134370.	99US-0145087.
PR 18-MAY-1999;	99US-0134768.	99US-0145089.
PR 19-MAY-1999;	99US-0134941.	99US-0145192.
PR 20-MAY-1999;	99US-0135124.	99US-0145145.
PR 21-MAY-1999;	99US-0135353.	99US-0145218.
PR 24-MAY-1999;	99US-0135629.	99US-0145224.
PR 25-MAY-1999;	99US-0136021.	99US-0145276.
PR 27-MAY-1999;	99US-0136392.	99US-0145913.
PR 28-MAY-1999;	99US-0136782.	99US-0145918.
PR 01-JUN-1999;	99US-0137222.	99US-0145919.
PR 03-JUN-1999;	99US-0137528.	99US-0145951.
PR 07-JUN-1999;	99US-0137724.	99US-0146386.
PR 08-JUN-1999;	99US-0138094.	99US-0146388.
PR 10-JUN-1999;	99US-0138340.	99US-0146389.
PR 14-JUN-1999;	99US-0139119.	99US-0147038.
PR 16-JUN-1999;	99US-0139452.	99US-0147204.
PR 17-JUN-1999;	99US-0139492.	99US-0147302.
PR 18-JUN-1999;	99US-0139454.	99US-0147303.
PR 18-JUN-1999;	99US-0139455.	99US-0147416.
PR 18-JUN-1999;	99US-0139456.	99US-0147493.
PR 18-JUN-1999;	99US-0139457.	99US-0147935.
PR 18-JUN-1999;	99US-0139458.	99US-0148171.
PR 18-JUN-1999;	99US-0139459.	99US-0148319.
PR 18-JUN-1999;	99US-0139460.	99US-0148565.
PR 18-JUN-1999;	99US-0139461.	99US-0148684.
PR 18-JUN-1999;	99US-0139462.	99US-0149175.
PR 18-JUN-1999;	99US-0139463.	99US-0149260.
PR 18-JUN-1999;	99US-0139750.	99US-0149722.
PR 18-JUN-1999;	99US-0139763.	99US-0149723.
PR 21-JUN-1999;	99US-0139817.	99US-0149929.
PR 22-JUN-1999;	99US-0139899.	99US-0149902.
PR 23-JUN-1999;	99US-0140353.	99US-0149930.
PR 23-JUN-1999;	99US-0140354.	99US-0150566.
		99US-0150884.
		99US-0151065.
		99US-0151066.
		99US-0151080.
		99US-0151303.
		99US-0151438.
		99US-0151930.

Db 34040 CCGACACGACACCAACGCCCACTCCACACACGACCCACACCCC 33997

RESULT 7
AAC44444
ID AAC44444 standard; DNA; 777 BP.
XX AC
AC AAC44444;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 42845.
XX DE Arabidopsis thaliana.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match					
C	1	100.8	3.5	390	4	US-09-197-649-7	Sequence 7, Appli
	2	80	2.8	11707	4	US-09-136-574A-1	Sequence 1, Appli
	3	79	2.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	4	71.2	2.5	494	4	US-09-056-556-176	Sequence 176, Appli
	5	70.2	2.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C	6	70.2	2.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	7	70	2.4	390	4	US-09-197-649-7	Sequence 7, Appli
C	8	68.8	2.4	18596	4	US-09-318-448-11	Sequence 11, Appli
	9	68.4	2.4	1931	2	US-09-130-114-2	Sequence 2, Appli
C	10	67.8	2.4	407	4	US-09-056-556-173	Sequence 2, Appli
	11	67.8	2.4	23673	4	US-09-773-816-1	Sequence 1, Appli
C	12	66.6	2.3	2293	1	US-08-604-913B-12	Sequence 12, Appli
	13	66.4	2.3	6416	4	US-09-136-574A-2	Sequence 2, Appli
C	14	66.2	2.3	1294	3	US-09-025-691-2	Sequence 2, Appli
	15	66	2.3	3004	1	US-08-276-213-6	Sequence 2, Appli
C	16	64.4	2.2	2064	1	US-08-343-428-1	Sequence 1, Appli
	17	64	2.2	985	4	US-09-056-556-182	Sequence 182, Appli
C	18	62	2.2	1280	4	US-09-060-756-4	Sequence 4, Appli
	19	61.8	2.2	6085	4	US-09-029-603-4	Sequence 4, Appli
C	20	61.6	2.1	1881	2	US-08-596-300A-4	Sequence 4, Appli
	21	61.6	2.1	3041	2	US-08-596-300A-5	Sequence 5, Appli
C	22	61.6	2.1	3718	2	US-08-596-300A-6	Sequence 6, Appli
	23	61.6	2.1	3796	2	US-08-596-300A-3	Sequence 3, Appli
C	24	61	2.1	1294	4	US-09-312-038-4	Sequence 3, Appli
	25	61	2.1	2289	4	US-09-312-038-3	Sequence 4, Appli
C	26	60	2.1	535	4	US-09-056-556-171	Sequence 171, Appli
	27	60	2.1	1439	4	US-09-056-556-167	Sequence 167, Appli


```

; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7

Query Match                2.4%;   Score 70;   DB 4;   Length 390;
Best Local Similarity      49.2%;   Pred. No. 7.3e-07;
Matches 184; Conservative 0; Mismatches 190; Indels 0; Gaps

Qy 2361 gccgtctccgtcggtgagtcgcgtcggtctgcgcgagcctgagcccgagcccgagcccgag 2420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 gccatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 62

Qy 2421 cagctcgccatccgcgtcgccgtcgccgagctgcagctgcattcctgcgtgcggtctccgtc 2480
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 122

Qy 2481 gccgtaccatcccgagtcggtctcggtctcggtccaccatcggtcgccgagcccgctc 2540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 182

Qy 2541 ttctgcaccgagccgctcttctgcacctcttcgtcgccgagcccaacgccgctgcgtc 2600
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 242

Qy 2601 gccggtctcgggtggggtgaaggtgcaggtatagaataatgatctcggcgcgggtgtgataa 2660
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 302

Qy 2661 tcagatcaagccgggtttgcaggtggtgataaccgggtcgctcggtggatttgtcgac 2720
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 362

Qy 2721 ggtgacggtgcggt 2734
      ||||| ||||| |||||
Db 363 gacgacgccaatggt 376

RESULT 8
US-09-318-448-11/c
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Query Match                2.4%;   Score 68.8;   DB 4;   Length 18596;
Best Local Similarity      61.9%;   Pred. No. 3.1e-06;
Matches 109; Conservative 0; Mismatches 67; Indels 0; Gaps

Qy 2428 ccatacgcgtcgccgtgcgcgagctcgtgcattccatcctcgtgcggtctcctgcgcgtc 2487
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10391 CCATCACCATCACCATCTCCATCTCCATCACCATCACCATCACCATCACCATCACCATCA 10333

```


2:

Best Local Similarity	34.3%	Fied. NO: 0.4e-00,
Matches	134; Conservative	0; Mismatches 113; Indels 0; Gaps 0;

y 2348 cqccgtccggatcggccgtctccgtcggtagtccggtcggcttcggcggagccctgagcccg 2407

RESULT 15
US-08-276-213-6

RESULT 14

IS-09-025-691-2

Sequence 2, Application US/09025691

PAGE NO. 8069259
GENERAL INFORMATION:

APPLICANT: Broadway, Roxanne M.

APPLICANT: Harman, Gary E.

TITLE OF INVENTION: FUNGUS AND CHITINOLYTIC

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON, HARGRAVE, D
STREET: CLINTON SQUARE P O

CITY: Rochester

STATE: New York

COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC DOS/MS D
SOFTWARE: PatentIn Release #1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/025
FILING DATE:

CLASSIFICATION:
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.
REGISTRATION NUMBER: 20 727

REGISTRATION NUMBER: 30,121
REFERENCE/DOCKET NUMBER: 1960

TELECOMMUNICATION INFORMATION:


```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-276-213-6

```

```

Query Match      2.3%; Score 66; DB 1; Length 3004;
Best Local Similarity 59.7%; Pred. No. 8.5e-06;
Matches 111; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 2429 catcgccgtcgccgtcgccgagctcgagtcacatcctcgctcgccgtctccgtcgccgtcac 2488
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2025 CGTCTGCATCGCCTAGCAGTCAACCGTCCCGTCCGTGTCGCCCTCTCCGTCCGCCGCC 2084
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 2489 catcgccgagtcgctctcggtctcctcgtaaccatcgccgtcgccgagcccgctcttcgtcac 2548
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2085 CGTCGGGAGTCCGACGCCGACGCTACTCCGAGCGCGACAGCCAGCCGCCGACGCCAACGC 2144
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 2549 cgagccgtcttcgtaccgtcttcgtcgccgagcccaagcccgctcgctcgccggtgt 2608
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2145 TGACCCCTACTGCTACGCCCCACGCCCGCAAGCCCGACGCCGCTACCCGACGCCGCT 2204
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 2609 cgggtg 2614
    ||| |||
Db 2205 CCGGAG 2210

```

Search completed: July 2, 2002, 16:53:17
Job time: 27704 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:33 ; Search time 2155.46 Seconds
(without alignments)
17964.965 Million cell updates/sec

Title: US-09-917-376-2

Perfect score: 2869

Sequence: 1 atggatcgttcggagaacat.....ggcggaacacacacacgacgn 2869

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_estt2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.4	3.3	559	12	BH306809
2	92.2	3.2	464	12	A2928926
3	90	3.1	546	10	BF072664
4	84.8	3.0	306	12	CNS07HOW
5	81.6	2.8	236	9	AA720369
6	81.2	2.8	478	10	BE496065
7	81.2	2.8	523	10	BF145280
8	81	2.8	1315	9	AV752548
9	80	2.8	565	12	AZ386138
10	79.6	2.8	873	9	BE217030
11	78.4	2.7	527	12	A2297141
12	75	2.6	1057	12	CNS033F1
13	74.6	2.6	401	10	BE499920
14	74.2	2.6	321	10	BG874723
15	72	2.5	535	10	BE494335
16	72	2.5	1208	9	AL514927
17	71.6	2.5	460	12	CNS0360U

18	71.6	2.5	522	9	AI601018
19	71.6	2.5	606	9	AI603748
20	71.2	2.5	587	10	C79507
21	71	2.5	259	12	A2374273
22	70.2	2.4	809	12	CNS02CFH
23	70	2.4	487	10	BE494661
24	70	2.4	490	10	BE494611
25	69.8	2.4	859	10	BF627594
26	69.2	2.4	439	9	AW496772
27	69.2	2.4	570	10	BG366070
28	69.2	2.4	696	9	AL506972
29	69.2	2.4	710	9	AL163439
30	68.6	2.4	365	12	L2644X
31	68.4	2.4	557	10	BF145396
32	68.2	2.4	279	9	AW496745
33	68.2	2.4	462	9	AU197994
34	68.2	2.4	623	10	BE495160
35	68	2.4	373	10	BF429394
36	67.8	2.4	584	10	BF255396
37	67.8	2.4	955	12	CNS02YA4
38	67.6	2.4	554	9	AW677991
39	67.2	2.3	427	10	BE494411
40	67.2	2.3	450	9	AU198202
41	67.2	2.3	829	12	CNS035GN
42	67	2.3	504	10	BM381428
43	67	2.3	651	10	BE419325
44	66.8	2.3	1009	12	CNS04065
45	66.6	2.3	458	10	BE587118

ALIGNMENTS

RESULT 1

BH306809

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH306809 559 bp DNA linear GSS 30-NOV-2001
CH230-100P7, TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-100P7, DNA sequence.
BH306809
GSS.
GI:17219217
Norway rat.
Rattus norvegicus
Craniata; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 559)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other_GSSs: CH230-100P7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering.information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 100 row: P column: 7
Seq primer: 17
Class: BAC ends.
Location/Qualifiers
1..559
/organism="Rattus norvegicus"
/strain="BN/SsNHsd/MCW"


```
Query Match          3.1%; Score 90; DB 10; Length 546;
Best Local Similarity 52.2%; Pred. No. 0.00024;
Matches 266; Conservative 0; Mismatches 220; Indels 24; Gaps 2;

QY 1266 gaagctcgctgagtgatgaagcagtgatgcaatcgatccggttcaactctgactgagtgct 1325
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GCACCTCGGCTGGATGATCGAATCCCTCGAGATCAACCCCTCGACAGCACCATTGGCT 111

QY 1326 ctacgaaacaggcgcagctgtgtacgcaacaaatgatctcaacgaagtgggactccgcgg 1385
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 CTACGGCACCGGCTCTCACTGTTTACGCGGCCACGACCTGACCAAGTGGACACCGTCCA 171

QY 1386 ccagattcatatcgcccgatgtcaaaagattggagagacgcggttaaacgatctcat 1445
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 CACGTGACCAATCCAGTCTTGCGCGCTTGGCATCGAAGAAATGGCTGTCTCGGTCTCGC 231

QY 1446 cagccgcgcgtctggcgcccgctcatcgagcgtctctcgagacctggcggttca---- 1501
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 CTCGGCCCCGGTGGCTCGGAGCTTCTCGCGGCGCTGCGTGACGACTGCGGCTTCACTT 291

QY 1502 -----ccacgcgcagcttactgcccgtgcacatcgacgatcttcacgtcaaccggtgtcac 1556
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 CAAGTCCAGTCCGACCTCGGCACCTCTCCCAAGACACCCCTGGATGACGCCCCAGTGGC 351

QY 1557 gaccgcaccagcgtcagctatcggaattgaatccgtcgatcatcgttcgcgtgaaag 1616
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 CAGCTCGGGAGCTGCTACGCGCGGCAACAAGCCGCCCAACGTCGTGCGCATCGG--- 408

QY 1617 ttctgatccatcgagcacaacccgacagcagcgctcgtcttcgacagacgcgcgca 1676
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 -----CTCGGCTCCGGCGCGCAGCAGGTGGCGGTATCTCTCGGACGGCGGCGC 456

QY 1677 gaactggttccaaggcagcgaacctggcgggggtgacagcggcgccacgctccgcacatc 1736
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 CTCGTGGCATGCTCACAAACGGCACCCACACCAACAGAGCAGCGGACGCTCGCCTACTC 516

QY 1737 ggcgcagcgtctcgtttcgtctggctcc 1766
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 TGCGGCGCGGACCACTATGCTGTGCTTC 546

RESULT 4
CNS07HOW
LOCUS
DEFINITION
  Anopheles gambiae GSS SP6 end of clone 24A06 of library NotreDame1
  from strain PEST of Anopheles gambiae (African malaria mosquito),
  genomic survey sequence.
ACCESSION
  AL610466
VERSION
  AL610466.1 GI:15916651
KEYWORDS
  GSS.
SOURCE
  African malaria mosquito.
  ORGANISM
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
    Culicoidae; Anopheles.
REFERENCE
  1 (bases 1 to 306)
  Genoscope.
  Direct Submission
  Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
REFERENCE
  2 (bases 1 to 306)
  Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H.
  Direct Submission
  Submitted (01-Oct-2001) BBMI, Institut Pasteur, 25, rue du Dr.
  Roux, Paris 75015, France
  This clone is from an A. gambiae BAC library provided by F.H.
  Collins and sequenced by Genoscope in collaboration with the
  Laboratory of Biochem. and Biol. Molec. of Insects, Institut
  Pasteur.
FEATURES
  source
    1..306
    /organism="Anopheles gambiae"

/strain="PEST"
/db_xref="taxon:7165"
/clone="24A06"
/clone_lib="NotreDame1"
/note="end : SP6"
BASE COUNT      65 a 110 c 63 g 68 t
ORIGIN

Query Match          3.0%; Score 84.8; DB 12; Length 306;
Best Local Similarity 70.6%; Pred. No. 0.0014;
Matches 113; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2425 tgcacatcgccgtcgccgtcgagctcgagtcacatctcgtcgccgtctcgtcgccg 2484
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCG 179

QY 2485 tcaccatcgccagtcgctcgtcgctcgctcaccatcgcgctcgcgagccgctcttcg 2544
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCG 239

QY 2545 tcaccgagccgctcttcgtcaccgcttcgttcgcccagcc 2584
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
240 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTTACCGTGCC 279

RESULT 5
AA720369
LOCUS
DEFINITION
  33562 Lambda-PRL2 Arabidopsis thaliana cDNA clone 171N14T7, mRNA
  sequence.
ACCESSION
  AA720369
VERSION
  AA720369.1 GI:2733979
KEYWORDS
  EST.
SOURCE
  thale cress.
  ORGANISM
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
  1 (bases 1 to 236)
  Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
  ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
  ,E. and Somerville,C.
  Genes galore: a summary of methods for accessing results from
  large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  Plant Physiol. 106, 1241-1255 (1994)
  95148729
  Contact: Thomas Newman
  MSU-DOE Plant Research Laboratory
  Michigan State University
  MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
  Lansing, MI
  Tel: 517-353-0854
  Fax: 517-353-9168
  Email: 22313tcn@bm.cl.msu.edu
  Seq primer: T7 dye primer.
FEATURES
  source
    1..236
    /organism="Arabidopsis thaliana"
    /strain="var columbia"
    /db_xref="taxon:3702"
    /clone="171N14T7"
    /clone_lib="Lambda-PRL2"
    /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
    Lambda PRL2 is a cDNA library derived from equal
    quantities of 4 pools of mRNA. The mRNA sources were 1) 7
    day germinated etiolated seedlings; 2) tissue culture
    grown roots; 3) staged plants half with 24 hour light
    cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
    same plants as 3 but aerial tissue (stems, flowers and
    siliques. The vector is BRL's lambda Zip-Lox. The cDNA
    inserts were directionally cloned with Sal-Not arms using
    oligo dT primed cDNA."
```



```

FEATURES
  source
    Location/Qualifiers
      1..1057
        organism="Tetraodon nigroviridis"
        db_xref="taxon:99883"
        clone="039109"
        clone_lib="A"
      /note="Genoscope sequence ID : COAA039CF05A1-end : T3"
BASE COUNT
285 a 214 c 329 g 207 t 22 others

```

	Query Match	2.6%	Score 75;	DB 12;	Length 1057;
	Best Local Similarity	67.7%;	Pred. No. 0.043;	Mismatches 105;	Conservative 0; Gaps 0
Qy	2426	cgcacatcgccgtcgcgctgcgcgagctcgaatgccatccgtctgcgcgtctcgcgtgcgcgt	2485		
Db	1030	CTCGGCTCTGTCTCCGTCTCTCCCTCCGTCCTCCGTCCTCCGTCCTCCTCGT	971		
Qy	2486	caccatcgccgagtccgtctcgtgttcctcaccatcggcgtcgcgcgagccgccttcttg	2545		
Db	970	CTCGCTCCGGTCTCGTCTCTGCTCTGCCCTCCGTCCTGCTCCCTCCGTCCTCGT	911		
Qy	2546	caccgagccgctttctgcataccgctttctgcgcg	2580		
Db	910	CTTGCTCTGTCTCGGTCCTCGGTCCTCGCTCCGTCCTCAG	876		

RESULT	13
LOCUS	BE499920
DEFINITION	WHE0976_F08_K16ZS Wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE0976_F08_K16, mRNA sequence.
ACCESSION	BE499920
VERSION	BE499920.1 GI:9698537
KEYWORDS	EST.
SOURCE	bread wheat.
ORGANISM	Triticum aestivum . Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 401)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES	SOURCE
Location/Qualifiers	
1. .401	
/organism="Triticum aestivum"	
/cultivar="Chinese Spring"	
/db_xref="taxon:4565"	
/clone="WHE0976.F08.K16"	
/clone_lib="Wheat pre-anthesis spike cDNA library"	
/tissue_type="Spike before anthesis"	
/dev_stage="Adult plant"	
/lab_host="E. coli SOLR"	
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site1: EcoRI; Site2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and	

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:39 ; Search time 101.03 Seconds
(without alignments)
813.566 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTPQYWSNVAIGGGFVD.....YGTNGRGIVYDGGAPSG 740

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	4.0	1684	AA19498	Bacterial amylase
2	159.5	4.0	1291	AAW59912	Amino acid sequenc
3	158.5	3.9	774	AAW35390	Flavobacterium odo
4	152.5	3.8	1751	AA113493	Truncated cellulase
5	152	3.8	745	AAAP90615	Polypeptide with i
6	152	3.8	3472	AAAP90913	Cenarchaeum symbio
7	151.5	3.8	776	AA190913	Isoamylase amino a
8	150.5	3.7	2367	AA198516	Novel human secret
9	148.5	3.7	750	AAU31850	Primary transcript
10	148	3.7	1045	ABG08332	Novel human diagno
11	148	3.7	1532	AAAB40945	Human ORFX ORF709

12	146	3.6	1749	22	AA19498	Glutamic acid rich
13	144.5	3.6	1472	22	AAW59423	Drosophila melanog
14	143.5	3.6	1532	22	AAU71956	Human bone marrow
15	143	3.5	894	15	AAAR47578	Flocculation prote
16	143	3.5	894	15	AAAR58754	S. cerevisiae PLO1
17	142.5	3.5	995	15	AAAR60812	Agarase 0107 from
18	142	3.5	925	22	AAAG92740	C glutamicum prote
19	141	3.5	754	21	AAAB35987	Sorbitol dehydroge
20	140.5	3.5	1000	21	AAAY58914	Fumonisin esterase
21	140.5	3.5	1000	21	AAAY68850	A fusion of aminop
22	140.5	3.5	1403	22	AAAB60540	Drosophila melanog
23	139	3.4	3118	22	AAU27790	Human full-length
24	137.5	3.4	1039	22	AAAB61734	Drosophila melanog
25	137.5	3.4	2464	22	AAAB63174	Drosophila melanog
26	137	3.4	911	22	AAU56103	Protonibacterium
27	136.5	3.4	1225	20	AAAY34538	Porphorymonas ging
28	136.5	3.4	1226	20	AAAY34537	Porphorymonas ging
29	136.5	3.4	1251	20	AAAY34404	P. gingivalis haem
30	136	3.4	2628	17	AAAR60300	Haemagglutinin pro
31	136	3.4	2628	19	AAAW69488	Fumonisin esterase
32	135.5	3.4	1205	21	AAAY58916	A fusion of aminop
33	135.5	3.4	1205	21	AAAY68852	Drosophila melanog
34	135	3.3	823	22	AAAB71484	Drosophila melanog
35	134.5	3.3	1805	22	AAAB65262	Mannanase. Cellul
36	134	3.3	1010	22	AAAB49442	Human polypeptide
37	133.5	3.3	530	22	AAO13658	Amino acid sequenc
38	133.5	3.3	1429	21	AAAY93606	Escherichia coli p
39	133.5	3.3	1778	22	AAAB52677	Banksia gouldi endo
40	133	3.3	875	19	AAAB34987	Drosophila melanog
41	133	3.3	1357	22	AAAB70775	Human wound healin
42	133	3.3	1712	22	AAAB44566	Human polypeptide
43	133	3.3	2058	22	AAAB97070	Soluble chitinase.
44	132.5	3.3	866	17	AAW02159	Vibrio furnissii e
45	132.5	3.3	866	21	AAAY52307	

ALIGNMENTS

RESULT 1
AA19498
ID AA19498 standard; Protein; 1684 AA.
AC AA19498;
XX 25-FEB-1992 (first entry)
DT Bacterial amylase A-180.
DE Maltopentose; G5; starch hydrolysis.
KW Maltopentose; G5; starch hydrolysis.
XX EP459385-A.
PN 04-DEC-1991.
XX 28-MAY-1991; 91EP-0108669.
XX 31-MAY-1990; 90DE-4017595.
XX (CONE) CONSOFT ELEKTROCHEM IND.
XX Schmid G, Candussio A, Bock A;
XX WPI; 1991-355676/49.
DR N-PSDB; AAQ13939.
PT New bacterial amylase, A-180 for malto:pentose prodn. - by
PT hydrolysis of starch, providing high yield and modifiable for
PT secretion from host cells
XX Disclosure; Page 7; 21pp; German.
XX The amino acid sequence is that of bacterial amylase A-180 which is

QY 180 TT---GVQSDIOGVV-----WVAFDKSSSLGQASKTIFVGVADPNPNVFWSDGGA 228
Db 396 ltlldlganynitsivliklpssiaaartqtiqlgh-----dqnttffsnlvsak 445
QY 229 TWAQVPCA-PTGIPHKGVFDPVNNHLYIATSNGTGPGYDSSGDVWKF-----275
Db 446 sysfdpasntvtip--vratvkrlqlnitsnsgap---agvvaefqvgtpapnptl 498
QY 276 SVTSGTWTTRISPVSDTDANDYFGYSGGLID---RQHPNFMVATOISWMPDIIIFRSTD 332
Db 499 tiftgmswsspvy-etd-----aitlnatvknngnasaaatvtnflynlnelagsap 548
QY 333 GGATWTRINDWTSYPNRSLRYVLIDIAEPWLTFGVQPNPPVSPKLGWDEMAIDPFNS 392
Db 549 vaal-----aagasaatvpinvgaaktaatyavgak-----vdesnavielne 599
QY 393 DRMLYGTGATL---YATNDL---TKW-----DSGGQIHIA 421
Db 590 snnsytnpaslvvapyvsssdlygtvstptpiannavsfvnlknqgtiasagghgvt 649
QY 422 PMYKGLEETAVNDLISPPSG--APLISA-----LGLDGGFT-----HADVTAVP---463
Db 650 vviknasgstvtqfsgsygslagpasvnltpgtwtaaagsyvtvatvaadanelpiqk 709
QY 464 ----STIFTSPTVTTCTSDYAEALNP-----SIIVRAGSFD----PSSQPNDRHVAF 507
Db 710 annantasiatvysargaspysrydtedatlgggtatksaptfdqaltaseatqglyaal 769
QY 508 STDGG-KNMF--OGSEPGGVTTGGTVAASADGSRFVWAPGDPQPVVYAVGFG-----557
Db 770 psgsyqlqwtvrgqggagvmtfmpdsadgm-----glngsladvvngtkvktvsilt 823
QY 558 --NSWAASQG-VPANAQ-----IRSDRVNPKTFYALSNGTFFYR-----STDGGVTF- 600
Db 824 syyswyfsgdmgdpapsagrpfrfdevhkwldtplkpgdtiricknngdsleygvdfi 883
QY 601 --QPVAAGL--PSSGAVGMFHAVP--GRE-----GDLWLAASGLYHSTN 640
Db 884 eiepvpaaisrpansvsvtdygvapndgqddltafkaavnaavasdkilypigtflh-- 941
QY 641 GGSWSAITGVSSAV-----NVGFGKSAPGSSYPAVFVVGTIG-----678
Db 942 -gmwe-igsvsmidhittagviytniqtfnanpasggislrirtgklfnsynvlnsn 999
QY 679 -----GVTGAIRS--DDCGT-----TWVLINDDQHOYGNWGOAITGDHAN 716
Db 1000 lrsryggnavykfgmdnfgtgnsvirdvwv----ehfecgfw----vgdygh 1042

RESULT 3

AAW35390
ID AAW35390 standard; Protein; 774 AA.

XX AC AAW35390;

XX DT 09-MAR-1998 (first entry)

XX DE Flavobacterium odoratum isoamylase.
XX KW Isoamylase; industrial enzyme.

XX OS Flavobacterium odoratum.

XX FH Key Location/Qualifiers
FT Peptide 1..33
FT Peptide /label= sig_peptide
FT Peptide 34..774
FT Peptide /label= mat_peptide

XX JP08023981-A.

XX PN 30-JAN-1996.

XX PD

XX 20-JUL-1994; 94JP-0167267.
XX 20-JUL-1994; 94JP-0167267.
XX (NISO) NIPPON SHOKUHN KAKO KK.
XX WPI: 1998-021932/03.
XX N-PSDB; AAT95400.
XX Flavobacterium odoratum isoamylase gene - useful in industrial processes
XX PS Claim 1; Page 6; 8pp; Japanese.
XX CC The present sequence is Flavobacterium odoratum isoamylase, which can be used as an industrial enzyme.
XX SQ Sequence 774 AA;

Query Match 3.9%; Score 158.5; DB 19; Length 774;
Best Local Similarity 18.8%; Pred. No. 0.016;
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;

QY 76 INTNKTWAA-----VGNMTNSWDPNDGAILRSSDOGATWITPL 114
Db 35 lnpnkigaaydatkanvtfkvyskatrielylylstatsaekakymtnsgslwsvt-i 93
QY 115 PFKLGNGMGRGRLAVLDPNNDNILYFGAPSGKGLWRSTDGATWSOMTNPDPVGTI 174
Db 94 p---tstlsgqglg-----gtlyygyrawgnpw---pynaswtkgssl-----130
QY 175 ANPTDTTGYQSDIQG-----VWVAFDKSSSLGQASKTIFVGVADPNNP 219
Db 131 -----gfisdvaagnrfnpnkllsdpyalelshdpttat-----tngs 170
QY 220 VFWSRDGGATWQAV---PGAPTGFI-----PHKGVFDPV---NHLVLIASNTG 262
Db 171 Iyas---gatyridngssapkgivlagdtqatgkptralkddvvyeahvrgltmndts 227
QY 263 -----GPYDSSGDVWKFSTVSGTWTRISPV-----PSTDTANDYFGYSGLT 305
Db 228 itaayrgtykgaalkaaalaaagvtaiefipvqetqndtndndpsstsgdnygymtlny 287
QY 306 ---DRO-----HPNTIMVATOISWMPDIIIFRSTDGATWTRIW 341
Db 288 fapdrtyaydktpgpgptrefkemvkafhdnngikvlvd-----vvynthtgeggawspt- 339
QY 342 DWTSYFNRSRLR-----YVLDISAE-PWLTFCGVQPNPPVPS-----PKLGWDEAM 395
Db 340 dkttnyitfrgldnptyysltadfqnswdntgvggnyntnrtiaqlnlvdsiaywrkl 399
QY 386 AIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLIS--PP---439
Db 400 gvdgyrfdlasvlgnsqchqcfmfkmdag-----nalnriavaelpat 445
QY 440 --SCAPLIS---ALG-----DLGCF-----THADVTAVPSTIFTSPVTTGTSDVYA 481
Db 446 gsgsvdliaepwaignsygvvgfpgsawaengayrdvvrqapnklgsvaitg-----499
QY 482 ELNPSIIVR-AGSFDSPSSQPNDRHVAFTDGGKNWFQG---SEPGGVTTGGTVAASADGS 537
Db 500 ----qmatrfagssd-----lygddgrkphwsnftahdgtfkldklyscsnkn 545
QY 538 RFVW--APGDGPQPVVYAVGFGNSW-----AASQ-----GVP-----ANAQ 571
Db 546 nqvwpypgsdggg-----dnnswdqggiaqdqrkaarnngmalmmlsaqvniavgdea 599
QY 572 IRS-----DRVNPKTFF-----ALSNGTFFRSTD-G 596
Db 600 lrsamcnnpnldssanlwnwrttdqnnfqsfkamiafrkahpalpanfyssvdsn 659

QY 597 GVTFQPVAAGLPSSGAVGVNF-----HAVPGK-EGDLWLAASSGLYHSTNGSSWSAIT 649
 Db 660 gnmeglrwfkpgdgvadatyfdannhaiaawridsefgdtasaiyvahn-----awsa-- 714
 QY 650 GVSSAVNVGFGKAPGSSYPFAVFGTIGVGTGAYRSDDCGTTWLVINDDQHOYGNWQ 708
 Db 715 -----qvnftlpwpgagkswrvtdtcgwaegasqvapgs-alvggentaygicgr 766

RESULT 4
 AAY13493
 ID AAY13493 standard; Protein; 1751 AA.
 AC AAY13493;
 XX 30-JUL-1999 (first entry)
 XX Truncated cellulases comprising amino acid sequence.
 DE Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing.
 XX Unidentified.
 OS EP921188-A2.
 PN 09-JUN-1999.
 XX 15-SEP-1998; 98EP-0810919.
 XX 19-SEP-1997; 97US-0932571.
 XX (CLRN) CLARIANT FINANCE BVI LTD.
 PA Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 PI Gibbs MD, Morgan H, Williams DP;
 DR WPI; 1999-315403/27.
 DR N-PSDB; AAX55662.
 XX New truncated cellulase proteins, useful in detergents and for
 PT producing 'stonewashed' denim
 XX Claim 7; Page 37-41; 65pp; English.
 PS The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel B6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.
 XX Sequence 1751 AA;
 SQ

Query Match 3.8%; Score 152.5; DB 20; Length 1751;
 Best Local Similarity 18.7%; Pred. No. 0.12;
 Matches 184; Conservative 101; Mismatches 349; Indels 351; Gaps 49;
 QY 6 YTWSNVAIGGGFVDGIVFNEGAPGILYVRTD-----IGMYRW-----D 45

Db 233 ylaatgyfytsggidlgw---aavwlyiatndssyltkaeelmseyangtntwtqcwdd 289
 QY 46 AANGRIWIPLDVGVNWNWNGYNGVVSIAADPINTKNKWAAGVM-YTNSWPDNDGAIIRSSD 104
 Db 290 vrygtlmlakitgkel--ykgavernldhwtidrtypkgnmayltgw---gslyrat 343
 QY 105 QG-----ATWQ-----ITPLPFKLGNMPCR-----GMGERLAVDPNDNDI 140
 Db 344 aafiacvyadwsgcdsnkktkylnfaksqidyalgst--grsfvvgfgntypqphnhrn- 400
 QY 141 LYFGAPSGKLMRSTDSGATWSQMTNFPD-----VGTYIANPTDTTGYQSDIQGVVW-- 192
 Db 401 -----ahsswansmkipeyhrhilygalvvgggsdsyndditdyvqne 444
 QY 193 VAFDKSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPCAPGTFPHKGVFPDVNH 252
 Db 445 vacydnagivgalak-----myqlyggepid--dfkaletptnd 481
 QY 253 VLYIAT--SNTGGPYDGSQDVWKFVSTGTTTRISPVSTDTAN-DYF-----GY 300
 Db 482 eifveskfgnsgqp---nytevislyyntgw---pprvtdklskfyldltellqagy 534
 QY 301 SGLTIDRQHPNTIMVATQISWMPDTTIFRSTDCGATWTRIMWTSYPNLSRLRYL-DISA 359
 Db 535 s-----pdvvkvd-----yyleggklsqgyvwd---knriyyvlvdfsg 572
 QY 360 EPWLTFGVOPNPVPSPKLGW---MDEAMAIDPFNSDRLYGTATLYATNDLTLTWDSG 415
 Db 573 tk-----iypggevehkkaqfkisvpggywdptn-dpsykgltsqleknkyaiaaydn 626
 QY 416 GOIHIAPMVKGLEETAVNDLISP-----PSGAP-----443
 Db 627 n-----lvwglepgaatstaptstptptptvtatptptptptptptptptptptpt 680
 QY 444 -----LISALGDL-----GGTHADVTAV-----P 463
 Db 681 vlyknnetsastgsirpfwkiwngssvdlrsvkirywytvdgkpkpsaavcdwaqlgas 740
 QY 464 STIFTSPTVFTTGS-VDYAELNPSIIVRAGSFPDSSQPNDRHVAFTDGGKN-----514
 Db 741 nvtfnfvklssgsvsgady-ylevgfsagagqlqpgkdtgdigvrfnkndwsynqaddds 799
 QY 515 WFQ-----GSEPGGVTTGGTVAASADGSRFVWAPGDP-----546
 Db 800 wlqsmtnygenakvtlylvgdvlwvgqepggaatpaptstatptptatptptptptptptpt 859
 QY 547 -----GQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSN-GT-- 589
 Db 860 vsatptpaptaspvg-gsywtbpsesygalkvwyangnlssptnvlqpk--ikienvgtta 916
 QY 590 -----FYRSTDCGVN-----FQP-----VAAGLPSSGA-----VG 614
 Db 917 vdlrsrvkvrywytiddeatqsvsvassinpaiyidvfkvlganaggadyveigfksag 976
 QY 615 VMFHAVPGREGDILWLAASSGLYHSTNGSSWSA-----ITGVSSAVNVGFGKAPGS 666
 Db 977 vlaagstkeirlsickqsgsynqndysvrsansyienekvtgyiddvlvlgregrna 1036
 QY 667 STPAVFWVTIGVGTGA----YRSDCCGT-----WVLINDDQHOYGNWQAAT 711
 Db 1037 qikvwyangnlgsmtavlnpklkienvgttavdlrsvkrywytiddeatqsvsvtsin 1096
 QY 712 GGHANLRRVYIGTNGRGI-VYGDIG 735
 Db 1097 payidvkvfkiganaggadyveig 1121

RESULT 5
 AAP90615
 ID AAP90615 standard; protein; 745 AA.
 XX
 AC AAP90615;

XX 03-NOV-1989 (first entry).
 XX Polypeptide with isoamylase activity from *Pseudomonas amyloclavata*.
 XX *Pseudomonas amyloclavata*; isoamylase; starch; malt;
 XX amyloclavate materials; maltose.
 XX *Pseudomonas amyloclavata*.
 OS EP327391-A.
 XX 09-AUG-1989.
 XX 03-FEB-1989; 89EP-0301088.
 XX 04-FEB-1988; 88JP-0024762.
 XX (HAYB) HAYASHIBARA SEIBUTS.
 XX Anemura A, Futai M;
 PI WPI; 1989-229398/32.
 DR N-PSDB; AAN90395.
 XX New polypeptide with isoamylase activity for converting starch to malt
 PT - is derived from *Pseudomonas amyloclavata* or transformed host
 PT microorganisms.
 XX
 XX Claim 5; page 22; 29pp; English.
 XX Polypeptide (see AAN90395) with isoamylase activity. The DNA is
 CC obt'd. from *Pseudomonas amyloclavata*, pref. strain SB-15, and expressed
 CC recombinantly. The polypeptide has mol. wt. 75-85 kD and isoelectric
 CC point 4.6-4.8. It is used, together with beta-amylase, to convert
 CC amyloclavate materials to high maltose prods. from which high purity
 CC maltose can be recovered. See AAP90610, AAN90390 for signal peptide.
 XX
 XX Sequence 745 AA;
 SQ
 Query Match 3.8%; Score 152; DB 10; Length 745;
 Best Local Similarity 19.6%; Pred. No. 0.042;
 Matches 160; Conservative 91; Mismatches 273; Indels 292; Gaps 43;
 QY 51 WIFLDDWGNWNGYN---GVVS---IAADPINTKVAAGWMTNSWDPNDGAILRSSD 104
 DB 82 wgp-nwpyasnwkgsgagcvadvadngdrfnknkllldpyaqevsqdp----- 129
 QY 105 QGATWQITPLPFKLGNNPGRGRLAVDPNDNDILYFGAPSGKGLWRSTDSGATWSQM 164
 DB 130 -----lpsnqn-----gnvfashyrttdsgl----- 152
 QY 165 TNFPDVGTYIANPTDTG-----YQSDIQGVVWVAFDKSSSLG----- 203
 DB 153 --yapkgvllvpstgtstgktptraqddviyevhrg-----fteqdtspaqyrgtyyg 205
 QY 204 --QASKTFVGVA-----DPNNPV-----FWSRGGGATWQA 232
 DB 206 aglkasylaslgvafeflpvqetqndandvpsndangnywmytenyfspdryaynk 265
 QY 233 VPCAPGT-FIPHGKVPDPNVHLYIATNTGGPYDGGSDGVWKFVTSCTGWTTRISVPST 291
 DB 266 aaggtcaefqamvqafnagikymdv-----vynhtaeggtwtssdpttat 312
 QY 292 -----DTANDYF-GYSLGATIDRQHPNTI---MVATQISWMPDPIIFRST 332
 DB 313 iyswrgldnatyyellsgnyfydnfganfnntvtaqnlivdsavaywntm---gvd 369
 QY 333 GGATWTRIDWTSYPRNSRYLVDISAEPLWTFGVQPNPVPSPKLGWMDMAIDPFNS 392
 DB 370 gfr-----fdlasvlgnsclnavhasa-----pncpngygnfdadnsnvai----- 410

QY 393 DRMLY-----GTGATLY-----ATNDLTWK-DSGGQIHIAPVMKLEETAVNDL 435
 DB 411 nrilreftvrpaaggtvclnrlngpsaatr--tsvdsrrvrvvsvyprqlrqagnel 468
 QY 436 ISPPSGAPLISALGDLGGTHADVTAVPTFTFTSPVFTTGTSDVYAEALNPSIIVR----- 490
 DB 469 -----gsmtiyvtqandfsgs-----snlfqsgsrspwnsinfidvhdgmltkdvysc 517
 QY 491 --AGSFPSSQPNDRHRAVAFSTGGKNWFGQSEPGVGTGTGTV---AASADGSRF-VWAPG 544
 DB 518 ngannsqasygppsdg---gtstnyswdqgmaag---tgaavdqrtaartgmafelmsag 570
 QY 545 DP---GQPVYAVGFGN-----SWAASQG---VPAQAQIRSDR---VNPX 580
 DB 571 tplmgggdeylrtlqcnmaynlldssanwltsywttdqsnfyfadrllrsarhlprps 630
 QY 581 TFVALSNGTFYRSTDCGVTFQPVAAAGLPSGAVGVNFHAPVPGKEGDLWLAAS-GLYHST 639
 DB 631 swysgqltwy-----qp-----sgava-----dsnywnntsyaiaai 665
 QY 640 NGSSWSAITGVSSAVNVGFGKSAPGVFVGTIGVGTGAYRSDDCGTTWVLIND 699
 DB 666 ngps-----lgdsnslyvayn-----gwsssvtftlpappsgtqwyrvtdt 706
 QY 700 QHOYGNWGAITGDHANLRRVYIGTNGRGIVYGDIG 735
 DB 707 -----cdwndgastfvapgssetlig--gagttgqcg 736
 RESULT 6
 AAY90913
 ID AAY90913 standard; Protein; 3472 AA.
 XX AAY90913;
 XX 30-AUG-2000 (first entry)
 XX Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO.4.
 XX Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
 KW characterisation; archae; therapeutic; industrial; laboratory.
 XX Cenarchaeum symbiosum.
 OS WO200018909-A2.
 PN 06-APR-2000.
 PD 29-SEP-1999; 99WO-US22752.
 PF 29-SEP-1998; 98US-0102294.
 PR (DIVE-) DIVERSA CORP.
 PA Swanson RV, Feldman RA, Schleper C;
 PI WPI; 2000-293148/25.
 DR N-PSDB; AAA55188.
 XX New nucleic acids and proteins isolated from the non-thermophilic
 PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
 PT physiology of these archae and in therapeutic, industrial or laboratory
 PT techniques -
 XX Claim 26; Page 102-110; 210pp; English.
 XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
 CC and proteins isolated from the non-thermophilic crenarchaeote
 CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in
 CC the present invention are useful in characterizing the physiology of
 CC these archae and can be used in therapeutic, industrial or laboratory
 CC techniques. AAA55227 to AAA55260 represent promoter sequences from
 CC Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and


```
QY 241 IPHKGVDPVNVHLYIATSNLTGPGYDGGSDGVWVKFVSTSGTWTTRISPVSPST----- 291
Db 302 qanvqafhaagikvymdv-----ynhtaeggtwtssdqttatiyswrgldn 348
QY 292 -----DTANDYF-GYSGLTIDRQHNTI--MVATQISWPPDTIIFRSTDDGATWTRW 341
Db 349 atyveltsnqfydyntgngantvntvaqnllvdsldaywant--gvdgrf-----f 400
QY 342 DMTS-YPNRLRYVLDISAEPMLTFCVQPNPPVSPKLGWMDAMAID----PFNSDRML 396
Db 401 dlasvlgnsclngaytasa-----pocpgngynfdaadsnvainrillreftvrpaa 451
QY 397 YGTGATLYATNDLTWKDSGGQIH-IAPVWKGLEE-----TAVNDLISPPSCAPLI 445
Db 452 gsgslldfa-----epwaigsnayqlggfpgqwsenlfrdsllrqael-----gsmci 502
QY 446 SALGDLGGTTHADVTAPVTFITSPVFTGTSDVYAEALNPSIIVR-----AGS 493
Db 503 yviqandfsgs-----snlfqsgsrspwnsinfidvhdgntlkdvyscngannsqawp 556
QY 494 FDPSSQPNDRHVAFTDGGK----NWFQSEPGGVTTGTGV----AASADGSRFVWAPGDP 546
Db 557 ygps-----dggststnyswdqgmsag---tgaavdgrraartgmafeml--sa 599
QY 547 GOPVYVAVGFGNSAAASQGVANQAIRSDRVNPKTF-YALSNGTFFYRSTDDGVTFOFVAA 605
Db 600 gtpimgg---gdeylrtlcnmnaaynlldasanwltyswttdqsnfytfagrliafrkabh 656
QY 606 GL-PSSGAGVGVMFHAVPGKEGDLWLAASGLYHSTNG----GSSWSAIT--GVSSAVNVG 658
Db 657 alrpss-----wysgsltwyqpsgavadsnywnntsnayalain-- 697
QY 659 FGKAPGESSYPAVFVGTGGVYGAVRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLR 718
Db 698 -gpslqdsniywayngwssvfttlpappsgtqwyrvldt-----cdwndgastfvapgs 752
QY 719 RVVIGTNGRGIYVGDIG 735
Db 753 etlig--gagttgygcq 767

RESULT 8
AAU31850
ID AAU31850 standard; Protein; 2367 AA.
XX
AC AAU31850;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2341.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
```

```
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 519; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
XX Sequence 2367 AA;
QY 3 TQP-----YTWSNVAIGGGFVDGIVFNEGAPGLIYVTRDIGMYRWDAAANGRWIPLLDWV 58
Db 1318 tqpnihdywnngigved-----wlaeerwlkfrfqtmdgmwensvswa 1360
QY 59 GWNWNGYGVVSTADPINTINKYAAVGMVYN--SMDPNDGAILRSDSQGATQITPLPF 116
Db 1361 ddeneasigswgasdkagiirswavacdetsvkw-----agarae----- 1402
QY 117 KLGNNPGRGCMGERLAVDPNNDNILYFGAPSGKGLRST-D-SGATWSQMTNFFDVGTYIA 175
Db 1403 ----nvvgigtwarageqasgg--lwaggtsetgwagdkasggawtgaenqasgswa- 1455
QY 176 NPDTTGYQSDIQGVVWVAFDKSSSL--COASKTIFGVGADPNPNPVFWSRDCGATWQAV 233
Db 1456 ----lagndq--igelwaagqasdgswpgqgqasgvswvgeaiggs--wt---gaenqas 1504
QY 234 PGAPTGFIPHKGVDPVNVHLYIATSNLTGPGYDGGSS-----GDVWKF---SVTSGTWTRI 285
Db 1505 egswag--agagnmssvsywavvvdqagggswagtsdqsggskprfedqasggsa-- 1560
QY 286 SPVPSDTDANDYFYSGLTIDRQHPNTIMVATQISWPPDTIIFRSTDDGATWTRINDWTS 345
Db 1561 -----gagg-----qasgsgmlgpe-----dqssgrswadtadqas 1591
QY 346 YPNRLRYVLDISAEPMLTFCVQPNPPVSPKLGWMDAMAIDPFNSDRMLYGTGATLYA 405
Db 1592 ggsr-lghvdqssggaw-----agtidqs-----gggskprf 1622
QY 406 TNDLTK---W-----DSGQIHIAPMVKLEETAVN--DLISPPSCAPLISALGDL--GG 453
Db 1623 enqtteegswagaggqagggskvgpedqsgsrswansgdqi---sgglvlgvdqanggs 1679
QY 454 FTHADVTAVPSTIFTSPVFTTGTG--VDYAEALNPSII--VRAGSFPDSSQPNDRHVAFT 509
Db 1680 wtga---ghpasvvgpkpfedqvsggrswadareqvvgdsrlglrdqssgds---wagtg 1733
QY 510 DGCKNWF---QSEPGGVTTGGTVAASADGSRFVWAPG-----DPGQPVVYAVG 555
Db 1734 dqasgwicvcpqsgtnggswggasgqdvvggsr----pgptngssagdsdpsgsv----- 1784
QY 556 FGNSWAASQGVPANQAIRSDRVNPKTFYALSNGTFFY-----RSTDGGVT--FQPVAAGLPS 609
Db 1785 sgscwtgagavdqagcskpggfedqai-----gggfwpgagdqtgggrrpqsedqssgigs 1840
```


Db	759	qwtsasastpdtacaaanthgha-esteasgqtqtsepasgsrirtsagatctpsssg---	814
Qy	600	FQPVAAGLPSSGAVG-----MFIHAVFGKGDGLWLAASSGILYHSTNGSGSSWASITGVS	652
Db	815	-----asgtpsgsegistggettrfnsnprdshtqtgttellsasashgal-pvstgma	869
Qy	653	SAYNVFGKSAPOGSSYPANFVVGITIGGVTG	682
Db	870	ss1-----vpgtfhptlseastagrtpg	992
RESULT 11			
AAB40945			
ID	AAB40945 standard; Protein; 1532 AA.		
XX	AAB40945;		
XX			
DT	08-FEB-2001 (first entry)		
XX			
DE	Human ORFX ORF709 polypeptide sequence SEQ ID NO:1418.		
XX			
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;		
KW	vulnery; antipsoiatic; antiparkinsonian; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antihypoid;		
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW			

KW	antidnaemic; gene therapy; cancer; proliferative disorder; hypertension
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive.
XX	
XX	Homo sapiens.
XX	
XX	WO200058473-A2.
PN	
XX	
XX	05-OCT-2000.
PD	
XX	
XX	31-MAR-2000; 2000WO-US08621.
PF	
XX	
XX	31-MAR-1999; 99US-0127607.
PR	
PR	02-APR-1999; 99US-0127636.
PR	
PR	05-APR-1999; 99US-0127728.
PR	
PR	30-MAR-2000; 2000US-0540763.
XX	
XX	(CURA-) CURAGEN CORP.
PA	
XX	
XX	Shimkets RA, Leach M;
PI	
XX	
XX	WPI: 2000-602362/57.
DR	
DR	N-PSDB; AAC75154.
XX	
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease .
XX	
XX	
PS	Claim 11; Page 1203-1206; 5507pp; English.
PS	
XX	
CC	AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	antithyroid; and antianaemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The

Db 279 glavqngsv-----ctngtnalvspilpgsvltisgnvdasgksniissggqvaiaqgdiav 334
QY 322 WPTIIFRSTGDGATWRIWMTSYPNRLRYVLDISAEPWLTFGVQPNPPVSPKLGWM 381
Db 335 qgatvdvnggggt-----vriggdfggq--lt-----lpnasqtli 370
QY 382 DEAMATDPNSDRMLYGTGATLYA--TNDLTKWDS-----GGQIHAPWVKGLEET-- 430
Db 371 dsnsvy---kadalltngngtviwaddstrfsgnisagggmtggn-----ggfvets 420
QY 431 -AVNDLISPPSGAPLISALCDLGCF-----THADVTAVPSTIFTSVPTTGTSTVD 479
Db 421 gakslnvddtarvntfatmgelgtwllpleiivgttdlldapkiv---svltittsid 477
QY 480 ----YAEALNPSIIVRAG--SFPDSSQPN-----DRHVAFTDGGKNNWFGQSEPGGVTT 526
Db 478 ngnvilqadqsiavqanfsadpsapgnltfdspitidalifslgtgsiifantgp--int 535
QY 527 GGTVAAS-----ADGSRFVWAGDPCQPVVYAVGFGNSWAASQGVPAQAQI-- 572
Db 536 gntlvtsfpntldfndkignanttftagp-----ydiyfrksvnggfdllignanfvy 588
QY 573 ---RSDRVNPKTYALNSNGFYRSTD---GGVTFOFVAAAGLP---SSGAVGVMEHAVP 621
Db 589 fddgagittlpkfsfvrateiyygndivtgnqifdgvyglqpnltssagsviftnni 648
QY 622 GKEGDLWLAASSGLYHSTNGSSWSAITGVSSAV-----NVGFCK-----SAPGS 666
Db 649 llngsiqvtaqniiv--sqpssslaveiasdvllnagqnsfgntrngnvdiaqlgn 706
QY 667 SYPAVFVGTIGGVTAIRSDDCGTTWVLINDQHQYGNWQQAITGDHANLRVYIGTN- 725
Db 707 istgsivtspfgnagnvniinaggt--lttgyietstngngdvttsagststaidtrg 764
QY 726 -GRGIVYGDIGGAPS 739
Db 765 fgdgleidsiggavs 779

RESULT 13
ID ABB59423
XX ABB59423 standard; Protein; 1472 AA.
XX ABB59423;
XX ABB59423;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 5061.
XX Drosophila melanogaster polypeptide SEQ ID NO 5061.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WIPI; 2001-656860/75.
DR N-PSDB; ABL03526.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX Disclosure; SEQ ID NO 5061; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1472 AA;
SQ

Query Match 3.5%; Score 144.5; DB 22; Length 1472;
Best Local Similarity 21.2%; Pred. No. 0.34;
Matches 191; Conservative 91; Mismatches 342; Indels 275; Gaps 45;

QY 14 GGGGFVDGIVFNEG-APGILYVRT-----DIGMYRWDAANGRW-----IP 53
Db 237 ggggrgggfvrsrgggrrmggrtggprgrgsgggpgay---gsrggnanedhheve 292
QY 54 LLD-----WGVNMMNGYGVVS-----TAADPINT 78
Db 293 lwdntiaaekqqaahddawgdwnneeyegskdkvfttsnatqsaanvsvgtgav 352
QY 79 NKVAAVGMVNTSWDP-----NDGAILRSSDQATWQITPLPKLGGNMPGRCMERL 131
Db 353 tavpaagaetisa--pbglehqlvggshleesssgpaavtp-patlisgattplldysa 410
QY 132 AVDPNNDNILYFCAPSGKGLWRSTD--GATWSQMTNE---PDVGTVIANTDTTGYQSD 186
Db 411 avsnpppqiqsgtqsgagtgasaaaggaggstpsfvsasptdfssaasaaatlhvhaq 470
QY 187 IQGVVWVAFD--KSSSLGQASKTIFGVADPNPNPFWSRDGGATWQAVP--GAPTGFIP 242
Db 471 kqqqlqqqttpikpsatlsveqsyfnsasq-----gvspsgvpsaapagyaq 520
QY 243 HKGVFDPVNHVLIATISNTGG--PYDSSGDVWKFVTSGTWT-----RI-- 285
Db 521 -----npv--aaysqtstsvgsqypntyanvfasgtaagtaeqsgqppqirrarvkl 573
QY 286 ---SPVPST-----DTANDYFGY-----SGLTIDRQHPNT 312
Db 574 ppskipasavempgdnalninnigylvqfgaldftddgfepipekvsgsfidgq--- 629
QY 313 IMVATQISWMPDTIIFRSTDGGATWRIWMTSYPNRLRYVLDISAEPWLTFGVQPNPP 372
Db 630 ----qgqqqpd-----dygksqq-----qqqvtlaaqlq--- 655
QY 373 VPSPKLGMDEMAAIDPFNSDRMLYGTGATLYATND--LTKWD-----SGQIHIAPMV 424
Db 656 --ssqisdalnaagytrstsqqqqgvssavnatldlqtsdpyggtgsgnaynqys 713
QY 425 KGLEETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPVFTTGTSDYAEIN 484
Db 714 sgaskta-----sgfpttap-----ggysstyanvqssvans-----yqqqq 751
QY 485 PSIIVRAGSFPDSSQPN-----DRHVAFTDGGKNNWFGQSEPGGVTTGTVAASADCSR 538
Db 752 -----ygsyqpsyyqqqagsgagsgtgavsggggtatqnpvgsssqnsgnassay 805
QY 539 FVWAPGDP-----GQPVYAVGFGNSWAASQGVPAQAQIIRSDRVNPKTYALNSNGFYR 592
Db 806 ltsygstpsqayqsgsvyngtglisn--sgfsgsasnas--sgyanfssasaklkdatas 862
QY 593 ST---DGVVTFQPVAAAGLPSSGAVGVMEHAVPVGKGDWLWLAASSGLYHSTNGGSSWAIT 649

XX AAR47578;
AC
XX 19-JUL-1994 (first entry)
XX
XX Flocculation protein of *Saccharomyces cerevisiae*.
XX
XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*;
KW flocculation protein; enzymatic process; fermentation;
KW biodegradation; catalysis.
XX
XX *Saccharomyces cerevisiae*.
XX W09401567-A.
XX
XX 20-JAN-1994.
XX
XX 07-JUL-1993; 93WO-EP01763.
XX
XX 08-JUL-1992; 92EP-0202080.
PR 14-DEC-1992; 92EP-0203899.
XX
XX (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
XX Klis FM, Schreuder MP, Toschka H, Verrips CT;
PI
XX WPI: 1994-035071/04.
DR N-PSDB; AAQ54029.
XX
XX Immobilisation of enzymes to microbial cell wall - by prodn. of
PT fusion protein of enzyme linked to anchoring protein
XX
XX Example 10; Page 59-64; 99pp; English.
XX
XX The flocculation protein is used in a method to immobilise enzymes
CC to a microbial cell wall. The coding sequence (FLO 1) is used in
CC the production of a recombinant polynucleotide which comprises a
CC structural gene encoding a protein with catalytic activity and at
CC least part of a gene encoding at least the C-terminus of a protein
CC capable of anchoring in a eukaryotic or prokaryotic cell wall. The
CC anchoring fragment or protein is selected from alpha agglutinin,
CC AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC proteinase of lactic acid bacteria. The recombinant polynucleotide
CC preferably also comprises a sequence encoding a signal peptide to
CC ensure secretion of the expressed product. The signal peptide is
CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.
CC The host microorganism can be used for performing enzymatic
XX processes on an industrial scale.
XX
XX Sequence 894 AA;
SQ

Query Match 3.5%; Score 143; DB 15; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.22;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGAPGILYVTRDGGMYRWDAANGRWIPLLDWVG----- 59
DB 53 tysnaaymaygya-----sktklgsvggtdtsidyn-----ipcvsssgtfpcpqe 99
QY 60 --WNNNGYNGVWSIAADPINTKNV--WAA--VGMYTNSWD----- 93
DB 100 dsgyngwckgmgaacs-----nsqgiaywstdlfgfyttnvltmgtgyflppqtsytfk 155
QY 94 ---PNDDAILRSSDQGATW-----QITPLPFKLGNNMPGRGMRGLAVDPNNDNIL 141
DB 156 fatvddsaill--svggatafnccagqppitstnftidgikpwgg-----slppniegtv 208

QY 142 YFGA-----PSGKGLMRSTDGATWSQMTNFPDPVGTGTYIANPTDTFTGYSQDIQGVVWVAFDX 197
DB 209 ymyagyyypm-kvysnavswgtlplsvtlpd-gttvsd-----dfegvyv-sfd- 255
QY 198 SSSSLGQASKTIFVGADPNPNVFWNSRDGATWQAVPGAPTGFIPHKGVDPDPVNHVLYIA 257
DB 256 --ddlsqsnct-----vpdpn-----yavsttttttpeptgtfststemtv 297
QY 258 TSNLTGGPYD-----GSSGDVWFSTGWTWTRISPVSTDTANDYFGYGLID-- 306
DB 298 tgnvgvptdetvivrptsegli-----sttpeptgtfststevit-igtngqptdet 353
QY 307 ----ROHPNTIMVATOISWMPDPIIFRSDGGATWTRIMDWTSYPNRSRYVL----- 355
DB 354 vivirptsegliettteptgtftstste-----mtvtgtnqgptdetvivrptseg 409
QY 356 --DISAEPWL-TFGVQPNPVPSPKLGWMDENAIIDPFNSDRMLYCTGATLYATNDLTWK 412
DB 410 lvttteptgttf-----tststemstvtgtnlptdetv-- 444
QY 413 DSGGQIHIAIPMVKGLEETAVNDLISPPSGA-----PLISALGDLGGFTHADVAV 462
DB 445 -----ivvktpttaisslsssssgqitssitssrpiitpfypsnq-tsvissv 493
QY 463 PSTIFTSPVTTGTSDVYAEINPSIIVRAGSFPDPSSQPNDRHVAFTDGGKNWFQSGEPG 522
DB 494 issvtssliftspvissvissstttstisfesskssvptssstsgsse-setssag 552
QY 523 GVTGTGTVAAASADGSRFVNAPCDGQPVVYAVGFGNSWAASQGVANAOIRSRVNPKTF 582
DB 553 svsssfissessksp---tyssslplvtssattsetassl-ppattktseqtlvt- 607
QY 583 YALSNGTFRYRSTDGGVTFQFVAAGLPSGAGVGMFHAVFG--KEGDLWLAASS----- 633
DB 608 -----vtsceshvctesispavstatv-----tvsgvtteyttwcpisttettkqt 654
QY 634 -GLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPAVF--VVGTTGGVTAIRSDDCG 690
DB 655 kgteqtdttktkgtvtvtvttisscesdvcsktas----pavtstatingvttey----- 704
QY 691 TTWVLIN 697
DB 705 ttwcpis 711

Search completed: July 2, 2002, 09:15:50
Job time: 238 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:46 ; Search time 39.87 Seconds
(without alignments)
453.347 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPYTWSNVAIGGGGFVD.....YIGTNGRGIVYGDIGGAPSG 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	162.5	4.0	772 2	US-08-410-784A-5
2	158.5	3.9	774 4	US-09-346-237-8
3	152.5	3.8	1751 4	US-09-136-574A-44
4	151.5	3.8	776 4	US-09-346-237-4
5	148.5	3.7	750 6	5457037-3
6	148.5	3.7	751 6	5457037-5
7	148.5	3.7	776 4	US-09-346-237-7
8	143	3.5	894 3	US-08-362-525-22
9	143	3.5	894 3	US-08-971-692-15
10	141	3.5	720 4	US-09-296-284-25
11	141	3.5	754 4	US-09-296-284-4
12	140.5	3.5	1000 4	US-09-352-159-25
13	140.5	3.5	1000 4	US-09-352-168-25
14	136	3.4	2628 2	US-08-570-311-14
15	135.5	3.4	1205 4	US-09-352-159-29
16	135.5	3.4	1205 4	US-09-352-168-29
17	132.5	3.3	626 6	5268290-2
18	132.5	3.3	866 1	US-08-386-727-8
19	132.5	3.3	866 2	US-08-600-452A-8
20	132.5	3.3	1395 4	US-09-540-245A-15
21	131.5	3.3	1026 2	US-08-542-003-6
22	131.5	3.3	1026 2	US-08-322-760A-6
23	130.5	3.2	862 1	US-08-325-267A-4
24	130.5	3.2	1326 4	US-09-147-236-5
25	130	3.2	1026 1	US-08-194-290-7
26	128	3.2	1042 3	US-08-928-361B-11
27	128	3.2	1837 3	US-08-928-361B-5

28	125.5	3.1	538 3	US-09-040-005-2	Sequence 2, Appli
29	125.5	3.1	538 4	US-09-522-217-115	Sequence 115, App
30	125.5	3.1	551 2	US-09-033-537A-1	Sequence 1, Appli
31	125.5	3.1	565 4	US-09-142-623-11	Sequence 11, Appl
32	125	3.1	735 2	US-08-313-185-48	Sequence 48, Appl
33	125	3.1	735 2	US-08-459-499-9	Sequence 9, Appli
34	125	3.1	735 2	US-08-459-499-12	Sequence 12, Appl
35	125	3.1	735 3	US-09-082-614A-48	Sequence 48, Appl
36	124	3.1	553 2	US-08-661-052-16	Sequence 16, Appl
37	124	3.1	553 4	US-09-188-082-16	Sequence 16, Appl
38	124	3.1	746 1	US-08-476-519-11	Sequence 11, Appl
39	124	3.1	746 5	PCT-US95-09323-11	Sequence 11, Appl
40	124	3.1	777 1	US-08-476-519-2	Sequence 2, Appli
41	124	3.1	777 5	PCT-US95-09323-2	Sequence 2, Appli
42	124	3.1	3031 1	US-07-689-008-2	Sequence 2, Appli
43	123	3.0	1160 3	US-08-808-599A-24	Sequence 24, Appl
44	122.5	3.0	802 4	US-09-147-236-4	Sequence 4, Appli
45	122.5	3.0	3519 4	US-09-428-517-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-410-784A-5
; Sequence 5, Application US/08410784A
; Patent No. 5912413
; GENERAL INFORMATION:
; APPLICANT: MYERS, ALAN M.
; APPLICANT: JAMES, MARTHA G.
; TITLE OF INVENTION: ISOLATION OF SU1, A STARCH DEBRANCHING
; TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
; TITLE OF INVENTION: SUGARY 1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,784A
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heine, Ph.D., Holliday C.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: ISU-002XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-2290
; TELEFAX: 617-451-0313
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-410-784A-5


```
Query Match      4.0%; Score 162.5; DB 2; Length 772;
Best Local Similarity 19.8%; Pred. No. 0.00071;
Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;

QY  51 WIPLLDWGNNNGYN---GWVS---IAADPINTNKVAAVGMVMTNSWDPNDGAILRSSD 104
D  108 WGP--NWPYASNGKSGACVSDVDANGDREPNKLLLDPPYAQESQDP-----155
QY  105 QGATWQITPLPFLKGNMGRGMRGLAVDPNDNDNLYFGAPSGKGLWRSTDSGATWSQM 164
D  156 -----LNPSNQ-----GNVFASAHRTDSCI-----178
QY  165 TNPEDVGTIANPTDTTG-----YQSDIOGVVWVAFDKSSSLG-----203
D  179 --YAPKGVVLVPSTQSTGCKPTRAQKDDVIYEVHVG-----FTEQDTSIPAQYRGTYG 231
QY  204 ---QASKTIFVGVA-----DPNNPV-----FWSRGGGATWQA 232
D  232 AGLKASYLSLGTAVEFLPVQETQNDANDVVPNSDANQNYGMYMTENYFSDRRYAYNK 291
QY  233 VPGAPTG-FIPKGVGDPVNVHLYXIATSTGGPYDSSGVDWKFESVTSCTWTTRISVPST 291
D  292 AAGGPTAEFOAMVQAFHNAKIKYMDV-----VYNHTAEGGTTWSSDPTTAT 338
QY  292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATOISWMPDTIIFRSTD 332
D  339 IYWRGLDNATYYELSGNOYFYDNTGIGANFNTYVTAQNLIVDSVAYWANTM---GVD 395
QY  333 GGATWTRIMWTYPNRSRLRYLIDSAEPWLTFGVQPNPPVPKLGWMDMAIDPFNS 392
D  396 GFR-----FDLASVLGNSCLNAVHASA-----PNCPNGGYNFDAAASNVAI-----436
QY  393 DRMLY-----GTGATLY-----ATNDLTKV-DSGQIHIAPMYKGLEETAVNDL 435
D  437 NRILRETVPAAAGTVWICLNLGPSAATR--TSWVDSRRVVRVEMSVPRQLRQONEL 494
QY  436 ISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPISIIVR-----490
D  495 ---GSMTIYVTDANDFSGS-----SNLFOSGSRSPWNSINFIDVHDGMLKDVYSC 543
QY  491 --AGSDPPSQPNDRHVAFSTGKKNWFQSEGGVTTGGTV---AASADGRF-VWAPG 544
D  544 NGANNSQASYPGSDG---GTSTNYSDWQMSAG---TGAADVQRRAAARTGMAFEMLSAG 596
QY  545 DP-----GQPVVYAVGFGN-----SWAASQG---VPANAQIRSDR---VNPX 590
D  597 TPLMOGGDEYLRILQCNNNAYNLDSSANWLTYSWTTDQSNFYTFQAQLIRSAHIPLRPS 656
QY  581 TFYALSNGTFTYRSTDGGVTQFPVAAAGLPSSGAGVGMFHAVPGKGLWLAAASSGLYHSTN 640
D  657 SWYSGSOLTWY-----QP-----SGAV-----ADSNWNNTS 683
QY  641 GGSWSAITGVSSAVNVGFKSAPGSSYPAVFVY---GTIGGVTVGAYRSDCGTTWLIND 698
D  684 NYAIAVAINGPSL-----GDSNDSIYVAYNGWSSSVTFLPAPPSGTQYRVTD 732
QY  699 DQHQYGNWQAITGDHANLRVRYIGYNGRIGVYGDIG 735
D  733 T-----CDWNDGASTFVAPGSETLIG--GAGTTYGQCG 763

RESULT 2
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629,200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
```


EARLIER FILING DATE: 1998-07-28
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 776

TYPE: PRT
ORGANISM: Pseudomonas amyloclavata
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)-(776)
OTHER INFORMATION: Isoamylase
US-09-346-237-4

Query Match 3.8%; Score 151.5; DB 4; Length 776;
Best Local Similarity 20.6%; Pred. No. 0.0045;
Matches 164; Conservative 91; Mismatches 329; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDVCGWNNWYGVVSIADPINTNKVWAAVG 86
DB 96 GITGAVY-----YGRANGPNWYASNWGKGSQAGFVSDVDANGDRPNKLLDPY 147
QY 87 MYTNSWDPNDAITLRSSDQATWQITPLPFLKLGNNPGRGMRGLAVDPNNDNILEFGAP 146
DB 148 AQEVSQDP-----LNPSNQ-----NGNVFASGASR-----TTDSGIY--AP 182
QY 147 SKGLWRSTDSGATWSQMTNFPDV--GTIANPTD--TTGQSDIQGVVWVAFDKSS--SS 201
DB 183 KGVVLPSTQSTGKTPRAQKDDVIYEVHVRGFTQDTSIPAOYRTYTGAGLKASYLAS 242
QY 202 LGQASTITFVGAD-----PN---NPVFW-----SRDGGATQAVPGAPT-G-F 240
DB 243 LG-VTAVEFLPQVETONDANDVVPNSDANQNYWGYMTENYFSPDRRYAYNKAAGGPTAEF 301
QY 241 IPHKGVPDPVNVHLYIATSTNGPYDSSGDVWKFVSTGTWTRISPVST-----291
DB 302 QAMVQAFHNAIKVYMDV-----VYNHTAEGGTWTSPTTATIYSRWGLDN 348
QY 292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWPPDTIIFRSTDGATWTRIW 341
DB 349 ATTYELTSGNQFYDNTGIGANFTYNTVAQNLIYDSLAWANTM---GVDGFR-----F 400
QY 342 DWT'S-YPNRSRLRVLDISAEPWLTFCGVPQNPVPSPKLGWMDMAID-----PFNSDRML 396
DB 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDAADSNVAINRLREFTVRPAA 451
QY 397 YGTGATLYATNDLTWKDSSGGQIH-IAPMVKGLFEE-----TAVNDLISPPSGAPLI 445
DB 452 GSGGLDLFA-----EPWAGGNSYQLGFFQGWSEWNGFLFRDSLRLQAQNEL-----GSMII 502
QY 446 SALGDLGGTHADVTAVPSTIFTSPVFTTGTSDVYAEALNPSIIVR-----AGS 493
DB 503 YVTDQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVIYSCGANNQAMP 556
QY 494 FDPSSQPNDRHVAFTDGGK----NWFQSEPGVITGGTV---AASADGSREVWAPGDP 546
DB 557 YGFS-----DGGTSTNYSWDQMSAG---TGAADQORRAARTGMAFEML--SA 599
QY 547 GQVWYAVGFGNSWAASQGVPPANAQIRSDRVNPKTE-YALSNGTFFYRSTDGGVTFOPVAA 605
DB 600 GTPLMQGG---GDEYLTLOCNNAYNLDSSANWLTYSWTDDQSNFTTFAQRLIAFRKAP 656
QY 606 GL-PSSGAVGVMPHVPKGEGLWLAAASGLYHSTNG-----GSSWSAIT--GVSSAVNNG 658
DB 657 ALRPS-----WYSGSLTWYQPSGAVADSNWNTSNYALAYAIN--697
QY 659 FGKAPGSSYPVAVVGTGTVGAVRSDCGTWTWVLIINDQHQYGNWGOAITGDHANLR 718
DB 698 -GPSLGDNSIYVAYNGWSSSVFTPLAPPSTQWRYRTDIT-----CDWMDGASTFVAPGS 752
QY 719 RVYIGTNGRCIVYGDIG 735
DB 753 ETLLIG--GAGTTYGQCG 767

RESULT 5
5457037-3

; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1.797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO: 3;
; LENGTH: 750
5457037-3

Query Match 3.7%; Score 148.5; DB 6; Length 750;
Best Local Similarity 20.6%; Pred. No. 0.0072;
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDVCGWNNWYGVVSIADPINTNKVWAAVG 86
DB 70 GITGAVY-----YGRANGPNWYASNWGKGSQAGFVSDVDANGDRPNKLLDPY 121
QY 87 MYTNSWDPNDAITLRSSDQATWQITPLPFLKLGNNPGRGMRGLAVDPNNDNILEFGAP 146
DB 122 AQEVSQDP-----LNPSNQ-----NGNVFASGASR-----TTDSGIY--AP 156
QY 147 SKGLWRSTDSGATWSQMTNFPDV--GTIANPTD--TTGQSDIQGVVWVAFDKSS--SS 201
DB 157 KGVVLPSTQSTGKTPRAQKDDVIYEVHVRGFTQDTSIPAOYRTYTGAGLKASYLAS 216
QY 202 LGQASTITFVGAD-----PN---NPVFW-----SRDGGATQAVPGAPT-G-F 240
DB 217 LG-VTAVEFLPQVETONDANDVVPNSDANQNYWGYMTENYFSPDRRYAYNKAAGGPTAEF 275
QY 241 IPHKGVPDPVNVHLYIATSTNGPYDSSGDVWKFVSTGTWTRISPVST-----291
DB 276 QAMVQAFHNAIKVYMDV-----VYNHTAEGGTWTSPTTATIYSRWGLDN 322
QY 292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWPPDTIIFRSTDGATWTRIW 341
DB 323 TTYELTSGNQFYDNTGIGANFTYNTVAQNLIYDSLAWANTM---GVDGFR-----F 374
QY 342 DWT'S-YPNRSRLRVLDISAEPWLTFCGVPQNPVPSPKLGWMDMAID-----PFNSDRML 396
DB 375 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDAADSNVAINRLREFTVRPAA 425
QY 397 YGTGATLYATNDLTWKDSSGGQIH-IAPMVKGLFEE-----TAVNDLISPPSGAPLI 445
DB 426 GSGGLDLFA-----EPWAGGNSYQLGFFQGWSEWNGFLFRDSLRLQAQNEL-----GSMII 476
QY 446 SALGDLGGTHADVTAVPSTIFTSPVFTTGTSDVYAEALNPSIIVR-----AGS 493
DB 477 YVTDQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVIYSCGANNQAMP 530
QY 494 FDPSSQPNDRHVAFTDGGK----NWFQSEPGVITGGTV---AASADGSREVWAPGDP 546
DB 531 YGFS-----DGGTSTNYSWDQMSAG---TGAADQORRAARTGMAFEML--SA 573
QY 547 GQVWYAVGFGNSWAASQGVPPANAQIRSDRVNPKTE-YALSNGTFFYRSTDGGVTFOPVAA 605
DB 574 GTPLMQGG---GDEYLTLOCNNAYNLDSSANWLTYSWTDDQSNFTTFAQRLIAFRKAP 630
QY 606 GL-PSSGAVGVMPHVPKGEGLWLAAASGLYHSTNG-----GSSWSAIT--GVSSAVNNG 658

Db 631 ALRPSS-----WYSGQLTWYQPSGAVADSNYNNWNTSYAIAIYAIN-- 671
QY 659 FGKSAFGSSPAVVFVVGTTGGVTGAYRSDCGTTWVLINDQHOYGNWGOAITGDHANLR 718
Db 672 -GPSLGDNSIYAYNGWSSVFTFLPAPPSGTQWYRVTD-----CDWNDGASTFVAPGS 726
QY 719 RVYIGTNGRGIYVDIG 735
Db 727 ETLIG--GAGTTYGCG 741
RESULT 6
5457037-5
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESE, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1.797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO: 5:
; LENGTH: 751
5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;
Best Local Similarity 20.6%; Pred. No. 0.0072;
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRDAANGRWIPLLDWYGNWNGYGVVSIADPINTNKVWAAVG 86
Db 71 GITGAVY-----YGYRAGPNWPYASNGWKGSGAGFSDVDANGDRFNPNKLLDYP 122
QY 87 MYTNSWDPNDAILRSSDQCATWQITPLPKLGGNPNPGRMGGERLAVDPNNDNILEYFAP 146
Db 123 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157
QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 158 KGVVLVPSTQSTGKTPTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYGAGLKASYLAS 217
QY 202 LGOAKTIFVGVAD-----PN---NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 218 LG-VTAVEELPVOETQNDANDVVPNSDANQNYGWTENYFSPDRRYAYNKAAGGPTAEF 276
QY 241 IPHKGVPDPNVHVLXIATSGTGPYDSSGDYKWFVTSCTWTRISPVSPST----- 291
Db 277 QAMVQAFHNAKIKVYMDV-----VYNHTAEGGTWISSDPTTIYSWRGLDN 323
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATQISWNPDTIIFRSTDCGATWTRIW 341
Db 324 TTYIELTSGNYFYDNTGIGANFNTYVAQNLIYDLSLAYWANTM---GVDGFR-----F 375
QY 342 DWTS-YPNRSLRVLDISAEPLTFGVQPNPPVPSPLGWMDMAID-----PFNSDRML 396
Db 376 DLASVILNGLANGAYTASA-----ENCNPGGYNFDADAASNVAINRILREFTVRPA 426
QY 397 YGTGATLYATNDLTKWDSGGQIHT-IAPMYKGLIE-----TAVNDLISPPSGAPLI 445
Db 427 GSGGLDLFA----EPWAGIGNSYQIGGFGQGWSEWNGLFDRSLRQONEL-----GSWTI 477
QY 446 SALGDLGGFTHADVATVPSTFIPTFTGTSYDVAELNPSIIVR-----AGS 493
Db 478 YVTQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVCYSGNANNQAMP 531
QY 494 FDPSSQPNDRHVAFSTDGK-----NWFQSGEPGGVTGTGTV---AASADGRFVWAPGDP 546

Db 532 YGPS-----DGGTSTNYSHWQCMASG---TGAAVDORRAARTGMFAEML--SA 574
QY 547 GQPVVYAVGFGNSWAASQGVPANAQIRSDRVNPKTF-YALSNGTFYRSTGGVTFQPVAA 605
Db 575 GTPLMQG---GDEYLRTLOCNNNAYNLDSSANWLTYSWTTDQSNFYTFQRLIAFRKAHP 631
QY 606 GL-PSSGAVGVMFHAVPGKEGLWLAASGLYHSTNG-----GSSWSAIT--GVSSAVNVG 658
Db 632 ALRPSS-----WYSGQLTWYQPSGAVADSNYNNWNTSYAIAIYAIN-- 672
QY 659 FGKSAFGSSPAVVFVVGTTGGVTGAYRSDCGTTWVLINDQHOYGNWGOAITGDHANLR 718
Db 673 -GPSLGDNSIYAYNGWSSVFTFLPAPPSGTQWYRVTD-----CDWNDGASTFVAPGS 727
QY 719 RVYIGTNGRGIYVDIG 735
Db 728 ETLIG--GAGTTYGCG 742
RESULT 7
US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

Query Match 3.7%; Score 148.5; DB 4; Length 776;
Best Local Similarity 20.6%; Pred. No. 0.0075;
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRDAANGRWIPLLDWYGNWNGYGVVSIADPINTNKVWAAVG 86
Db 96 GITGAVY-----YGYRAGPNWPYASNGWKGSGAGFSDVDANGDRFNPNKLLDYP 147
QY 87 MYTNSWDPNDAILRSSDQCATWQITPLPKLGGNPNPGRMGGERLAVDPNNDNILEYFAP 146
Db 148 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 183 KGVVLVPSTQSTGKTPTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYGAGLKASYLAS 242
QY 202 LGOAKTIFVGVAD-----PN---NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 243 LG-VTAVEELPVOETQNDANDVVPNSDANQNYGWTENYFSPDRRYAYNKAAGGPTAEF 301
QY 241 IPHKGVPDPNVHVLXIATSGTGPYDSSGDYKWFVTSCTWTRISPVSPST----- 291
Db 302 QAMVQAFHNAKIKVYMDV-----VYNHTAEGGTWISSDPTTIYSWRGLDN 348
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATQISWNPDTIIFRSTDCGATWTRIW 341


```

; 349 TTYELTSGNYFYDNTGIGANFNTYNTVAQNLIIVDSLAYWNTM---GVDGFR-----F 400
; 342 DWTS-YNRSRLRYVLDISAEPWLTFTGQPNPVPSPKLGWDEAMID-----PNSDRML 396
; 401 DLASVLGNSCLNAYTASA-----PNCPPGGYNFDDAASNVAIRILREFTVRPAA 451
; 397 YGTGATLYATNDLTWDSGGQIH-IAPMWKGLIE-----TAVNDLISPPSGAPLI 445
; 452 GSGLDLFA---EPWAIAGNSYQLGGFPQGWSEWNGSLFRDSLRLQAQNEL-----GSMTI 502
; 446 SALGDLGGFTHADVTAVPSTFIPTFTGTSVDYAEALNPSIIVR-----AGS 493
; 503 YVTQDANDFSGS---SNLFSSGSRSPWNSINFIDVHDMTLKDYISCGANNQAWP 556
; 494 FDPSSQPNDRHAFVSTDDGK---NMFQSEPGGVTTGGTV---AASADGSRFVWAPGDP 546
; 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVORRAARTGMAFEML---SA 599
; 547 GQPVVAVGFGNSWAASQGVYVANAQIRSDRVNPKTF-YALNSGTFTYRSTDDGGVTFQPVAA 605
; 600 GTPLMQG---GDEVLRTLQCNNAYNLDSSANWLTYSWTTDQSNFYTFAORLIAFRKAHP 656
; 606 GL-PSSGAVGVMPHVPCKEGDLWLAASSGLYHSTNG-----GSSWSAIT--GYSSAVNVG 658
; 657 ALRPSS-----WYSGSLTWYQPSGAVADSNIYWNNTSNYAIAYAIN-- 697
; 659 FGKSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLR 718
; 698 -GPSLGDNSIYVAYNGWSSSVFTLPAPSPGQWRYVDTT---CDWNDGASTFVAPGS 752
; 719 RVVIGNRGIVYGDIG 735
; 753 ETLIG--GAGTTYGQCQ 767

RESULT 8
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763

```

```

; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-525-22

Query Match 3.5%; Score 143; DB 3; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.023;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGAPGILYVRDIDGMYRWDAANGRWIPLLDWVG----- 59
Db 53 TYSNAAYMAYGYA-----SKTLGSGVGGQTDISIDYN-----IPCVSSTGTFPCPQE 99
QY 60 --WNNNGYGVVSIADPINTNKV--WAA--VGMVTNSWD----- 93
Db 100 DSYGNNGCKRGMGACS---NSQGIAYWSTDLFGFYTPTNVILEMTGYFLPQTGYSYTK 155
QY 94 ---PNDGAILRSDQCATW-----QITPLPFKLGGNMPGRGMYERLAVDPNNDNIL 141
Db 156 FATVDDSAIL--SVGGATAFNCCAQQPPITSTNFTIDGIKPMGG-----SLPPNIEGTV 208
QY 142 YFGA---PSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDITGYQSDIQGVVWAFDK 197
Db 209 YMTAGYYPM-KVVISNAVSWGLPISVLDP-GTVSD-----DEGVYI-SFD- 255
QY 198 SSSSLGQASKTIFGVADPNPNPVFWSRDGGATQWQVPGATGFIPIHKGVPDPVNHVLYIA 257
Db 256 --DDLQSQNT---VPDPN-----YAVSTTTTTEPTWTGTSTSTEMTV 297
QY 258 TSNITGGPYD-----GSSGDVWKFESVTSQWTRISVPVSTDTANDFYSGSLID-- 306
Db 298 TGTNGVPTDETIVIRTPTEGLI---STTEPTGTFTSTSTEVTT-ITGNGQPTDET 353
QY 307 ---RQHPNTIMVATQISWMPDIIIFRSTDDGATWRIWDWTSYPNKSLRYVL----- 355
Db 354 VIVIRPTSEGLISTTEPTWTGTFTSTSTB---MTVTGTNCGPTDETIVIRTPTESEG 409
QY 356 --DISAEPWL-TFGVQPNPPVPSFKLGWMDMAIDAIDPFNSDRMLYGTGATLYATNDLTKW 412
Db 410 LVTTTTEPTWTGT-----TSTSTEMSTVTGTNGLPTDETIV-- 444
QY 413 DSGGQTHIAPMWKGLEETAVNDLISPPSGA-----PLISALGDLGGFTHADVTAV 462
Db 445 -----IVVKTPTTAISSLSSSSQITSTSSRPIITFPYPSNG-TSVISSV 493
QY 463 PSTIFTSPVTTCTSDYAEALNPSIIVRAGSFPDPSQPNDRHVAFTDCKGNWFOQSEPG 522
Db 494 ISSSVTSSLIFTSSPVISSSVISSTTTSIFSESSKSSVIPTSSSTSSSE-SETSSAG 552
QY 523 GVTGTGTVAAADGSRFVWAPGDPQPVVYAVGVFGNSWAASQGVVANAQIRSDRVNPKTF 582
Db 553 SVSSSSFISSSESKSP---TYSSSLPLVTISATTSSETASSL-PPATTTKTSQTLT- 607
QY 583 YALNSGTFTYRSTDDGGVTFQPVAAAGLPSGAVGVMPHVPG--KEGDLWLAASS----- 633
Db 608 -----VTSCESHVCTESISPAIVSTATV-----TVSGVTTEYTTWCPISTETTQT 654
QY 634 -GLYHSTNGSSWSAITGYSSAVNVGFGKSAPGSSYPAVE--VVGITGGVTAIRSDDCG 690
Db 655 KGTTEQTETTKQTQTTVTVTSSCESDVCSKTAS-----PAIVSTSTATINGVTTEY----- 704

```


QY 691 TTWVLIN 697
| | | | |
Db 705 TTWCPIS 711

RESULT 9

US-08-971-692-15

; Sequence 15, Application US/08971692
; Patent No. 6114147
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immobilized proteins with specific binding
; capacities and their use in processes and products.
; NUMBER OF SEQUENCES: 40
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,692
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-971-692-15

Query Match 3.5%; Score 143; DB 3; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.023;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGACILVVRTDIGMYRWDAANGRWIPLLDWVG----- 59
| | | | | : : : : : | | | | | : : : : :
Db 53 TYSNAAIYAWGYA-----SKTKLGVGGQDTSIDYN-----IPCVSSSGTFPCQE 99
| | | | | : : : : : | | | | | : : : : :
QY 60 --WNNWNGYGVSIADPINTNKV--WAA--VGMYTNSWD----- 93
| | | | | : : : : : | | | | | : : : : :
Db 100 DSYGNWCKMGACGACS-----NSQGIAYWSTDLFGFYTTNTVLEMTGYELPQGTGSYTFK 155
| | | | | : : : : : | | | | | : : : : :
QY 94 --PNDGAILRSSDQATW-----QITPLPFLKGGNMPGRGMRGERLAVDPNNDNIL 141
| | | | | : : : : : | | | | | : : : : :
Db 156 FATVDDSAIL--SVGGATAFNCCAQQQPITSTNFTIDGKPMGG-----SLPPNIEGTV 208
| | | | | : : : : : | | | | | : : : : :
QY 142 YFGA----PSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDITGYQSDIQGVVWAFDK 197
| | | | | : : : : : | | | | | : : : : :
Db 209 YMYAGYYYPW-KVYYSNAVSWGILPISVTLPD-GTIVSD-----DEGYIY-SFD- 255
| | | | | : : : : : | | | | | : : : : :
QY 198 SSSSLGQASKTIFGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVDPVNVHVLIA 257
| | | | | : : : : : | | | | | : : : : :
Db 256 --DDLQSQNST--VDPDSN-----YAVSTTTTTEPWTGTSTSTEMTTV 297
| | | | | : : : : : | | | | | : : : : :
QY 258 TSNTGGPYD-----GSGDWKWFVSQGTWTRISPVPTDNDYFGYSLGLTID-- 306
| | | | | : : : : : | | | | | : : : : :
Db 298 TGINGVPTDETIVIRPTSEGLI---STTEPWTGFTTSTSEVTT-ITGNGQPTDET 353
| | | | | : : : : : | | | | | : : : : :
QY 307 ----ROHPNTIMVATQISWMPDIIIFRSTDGATWTRIDWTSYPNRSLRYVL----- 355
| | | | | : : : : : | | | | | : : : : :
Db 354 VIVIRPTSEGLISTTTEPWTGFTSTSE---MTVTGNTGQPTDETIVIRPTSEG 409
| | | | | : : : : : | | | | | : : : : :
QY 356 --DISAEPWL--TFCGVQPNPVPSPKLGWMDAIDAIDPFNSDRMLYGTGATLYATNDLTKW 412
| | | | | : : : : : | | | | | : : : : :
Db 410 LVITTTTEPWTGTF-----TSTSTENSTVGTGNTGLPTDETIV-- 444
| | | | | : : : : : | | | | | : : : : :
QY 413 DSGQIHIAPMVKLEETAVNDLISPSPA-----PLISALGDLGFGFHADVTAV 462
| | | | | : : : : : | | | | | : : : : :
Db 445 -----IVWKTPTTALSSLSSSSGQITSSITSSRPIITFPYPSNG-TSVISSV 493
| | | | | : : : : : | | | | | : : : : :

QY 463 PSTIFTSPIVETTGTSVDYAEINPSIIIVRAGSFDSPSSOPNDRHVAFTSDGKKNFQSGEPG 522
| | | | | : : : : : | | | | | : : : : :
Db 494 ISSSVTSSSLFTSPVSSSVISSSTTTTSTIFSESSKSSVIPTSSSTSGSSE-SETSSAG 552
| | | | | : : : : : | | | | | : : : : :
QY 523 GVTGTGTVAAASADGSRFVWAPGDPGVVYAVGFGNSWAAASQGVPAANAQIRSDRVNPKTF 582
| | | | | : : : : : | | | | | : : : : :
Db 553 SVSSSFISSESSKSP---TYSSSLPLVTSATTSQETASSL-PPATTTTSEQTILVT- 607
| | | | | : : : : : | | | | | : : : : :
QY 583 YALSNGTFFYRSTGQVTFQPVAAAGLPSSGAVGMFHAVPG--KEGDLWLAASS----- 633
| | | | | : : : : : | | | | | : : : : :
Db 608 -----VTSCESHVCTESISPAIVSTATV-----TVSGVTTEYTTWCPISTTETTKQT 654
| | | | | : : : : : | | | | | : : : : :
QY 634 -GLYHSTNGSSSAITGVSSAVNVNFGKSPGSSYPAVF--VVGTTGGVTGAYRSDDCG 690
| | | | | : : : : : | | | | | : : : : :
Db 655 KGTEQTETTKOTTVVTTISSCEDVCSTAS---PAIVSTSTATINGVTTEY----- 704
| | | | | : : : : : | | | | | : : : : :
QY 691 TTWVLIN 697
| | | | |
Db 705 TTWCPIS 711
| | | | |

RESULT 10

US-09-296-284-25
; Sequence 25, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-25

Query Match 3.5%; Score 141; DB 4; Length 720;
Best Local Similarity 20.4%; Pred. No. 0.024;
Matches 145; Conservative 73; Mismatches 235; Indels 258; Gaps 43;

QY 26 EGAP----GILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVV--SIA----ADP 75
| | | | | : : : : : | | | | | : : : : :
Db 58 EGTPLIVDGVMYATTNWSKMKALDAATGKLL-----WSYDPKVPNGIADRGCCDT 107
| | | | | : : : : : | | | | | : : : : :
QY 76 INTNKVWAAVGMVNTNSWDPNNDGAI--LRSSDQATWQITPLPKLGNMFGRMGERLAV 133
| | | | | : : : : : | | | | | : : : : :
Db 108 VNRGAAYWNGKVYFGFTG---DGRLLIALDAKTGLKLVMSVYTVP-----KEAOLGHQRSYTV 159
| | | | | : : : : : | | | | | : : : : :
QY 134 DPNNNDNILEYFAPS-GKGLWRSTDGATWSQMTNFPDVGTYIANPTDITGYQSDIQGVVW 192
| | | | | : : : : : | | | | | : : : : :
Db 160 D-----GAPRIAKGVIIIGNGAEGF-ARCF-----VTAYDAETGKMDW 197
| | | | | : : : : : | | | | | : : : : :
QY 193 VAF---DKSSSLGQASKTIFGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVDP 249
| | | | | : : : : : | | | | | : : : : :
Db 198 RFTVTPNPNKPDGAASDDVLMSKAYPT---WGK--GGAWKQQGGG--GTVWDSLIIYDP 249
| | | | | : : : : : | | | | | : : : : :
QY 250 VNHVLIATNTGPGYDSSGCDVWKFVSQGTWTRISPVSTDTANDYFGYSLGLTIDRQH 309
| | | | | : : : : : | | | | | : : : : :
Db 250 VTDLVLVGVGN-GSPWN-----YKFR-----SEKGNLFLGSIVAIN--- 286
| | | | | : : : : : | | | | | : : : : :
QY 310 PNTIMVATQISWMPDIIIFRSTDGATW-----TRIDWTSYV-----NRSURYV 354
| | | | | : : : : : | | | | | : : : : :
Db 287 -----PDV-----GKYVWHFQETPMDQWDITSVQOIMALDMPVNGEMRHV 326
| | | | | : : : : : | | | | | : : : : :
QY 355 LDISAEPWLTFGVQPNPVPSPKLGWMDAIDAIDPFNSDRMLYGTGATLYATNDLTKWDS 414
| | | | | : : : : : | | | | | : : : : :


```
Db 327 L-----VHAPKNGFF---YIIDA-KTGKFIISGKPYT-----YENWAN 359
QY 415 G-----GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDGLGFTHADVTAVPSTIFTS 469
Db 360 GLDPVTRGRPNYP-----DALWTLNGKWPYIGIPDGLGHNFAAMAYSEQ---TK 405
QY 470 PVFTTGTSDYAEALNPSIIVRAGSFDPSQPNDHRHVAFTSDGKKNWFQSGEPG--GVTTG 527
Db 406 LVYIPAQQVFPV-----YDPKGGFKAH-----HDSWNLGLDMKNKIGLLDD 446
QY 528 GTVAASADGSRFV-----WAPGDPGPQPVVYAVGFGNSWAASQGVPAANAQIRSDRVN 578
Db 447 NDPQHKADKAQFLKDLKGWIVAMPD--QKQAAFTVDHKGPM-----486
QY 579 PKTFYALNSNGTFYRSDTGGVTFQPVAAAGLPSSGAGVGMFHAVPGKEG-DLW-LAASGLY 636
Db 487 -----NGLL-ATAGGVLFQGLANG-----EFHAYDATTKDKLFTTFAQSAII 528
QY 637 -----HSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVVGTIGGVGTG 682
Db 529 APPVTYTANGKQ-----YVAVEGWM-----GGIYP--FFLGGVARTSG 564

RESULT 11
US-09-296-284-4
; Sequence 4, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-4

Query Match 3.5%; Score 141; DB 4; Length 754;
Best Local Similarity 20.4%; Pred. No. 0.026;
Matches 145; Conservative 73; Mismatches 235; Indels 258; Gaps 43;

QY 26 EGAP-----GILYVRTDIGMYRMDAANGRIPLLDWYGNNGYGVW--SIA-----ADP 75
Db 92 EGTPLIVDGYMYATTNWSKMKALDAATGKLL-----WSYDPKVPGNIDRGCDDT 141
QY 76 INTNKVAAVGMTNSWDNDGAI--LRSSDQAGTQWITPLPFKLGNNMPPGRMGERLAV 133
Db 142 VNREGAATWNGKVTFGTG---DGRLLALDAKTGKLVMSVTVTP-----KEAQLGHQRSYTV 193
QY 134 DPNNNDILYFGAPS-GKGLWRSTDGATWSQMTNEPDVGTYIANPTDTTCGYQSDIOGVVW 192
Db 194 D-----GAPRIAGKVIINGGAEEFG-ARGF-----VTAIDAEKGMKW 231
QY 193 VAF---DKSSSLGQASAKTIFGVADPNPNPVFWSRDGATQWAVPGAPTGFIPHKGVFDP 249
Db 232 REFTVTPNDKPDGAASDDVLMSKAYPT---WGK--GGAWKQGGG--GTWVDSLLIYDP 283
QY 250 VNHVLIATNTGPGVDGSSGDVWKSFSVTGWTTRISPVSTDTANDYFGYSGLTIDRQH 309
Db 284 VTDLVVLGVGN-GSPWN-----YKFR-----SEKGNMLFLGSIVAIN---320
QY 310 PNTIMVATQISWMPDRIIFRSTGGATW-----TRIDWMTSYN-----NRSIRVY 354
Db 321 -----PDT-----GKVVHFQETPMQWDITTSYQQIMALDMPVNGEMRHV 360
```

```
QY 355 LDISAEPWLTFGVQPNPPVSPKLGWMDAEMAIIDPNFSDRMILYGTGATLYATNDLTKWDS 414
Db 361 L-----VHAPKNGFF---YIIDA-KTGKFIISGKPYT-----YENWAN 393
QY 415 G-----GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDGLGFTHADVTAVPSTIFTS 469
Db 394 GLDPVTRGRPNYP-----DALWTLNGKWPYIGIPDGLGHNFAAMAYSEQ---TK 439
QY 470 PVFTTGTSDYAEALNPSIIVRAGSFDPSQPNDHRHVAFTSDGKKNWFQSGEPG--GVTTG 527
Db 440 LVYIPAQQVFPV-----YDPKGGFKAH-----HDSWNLGLDMKNKIGLLDD 480
QY 528 GTVAASADGSRFV-----WAPGDPGPQPVVYAVGFGNSWAASQGVPAANAQIRSDRVN 578
Db 481 NDPQHKADKAQFLKDLKGWIVAMPD--QKQAAFTVDHKGPM-----520
QY 579 PKTFYALNSNGTFYRSDTGGVTFQPVAAAGLPSSGAGVGMFHAVPGKEG-DLW-LAASGLY 636
Db 521 -----NGLL-ATAGGVLFQGLANG-----EFHAYDATTKDKLFTTFAQSAII 562
QY 637 -----HSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVVGTIGGVGTG 682
Db 563 APPVTYTANGKQ-----YVAVEGWM-----GGIYP--FFLGGVARTSG 598

RESULT 12
US-09-352-159-25
; Sequence 25, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 1000
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-25

Query Match 3.5%; Score 140.5; DB 4; Length 1000;
Best Local Similarity 20.9%; Pred. No. 0.041;
Matches 159; Conservative 69; Mismatches 224; Indels 309; Gaps 41;

QY 150 GLWRSTDGATWSQMTNEPDVGTYIANPTDTTCGYQSDIOGVVWVAFDKSSSLGQASAKTI 209
Db 16 GLSASLASGAPTAKI-----DAGMVVGTITTVPGTTATVSEFLGVFP-----AASPTR 63
QY 210 FVGVADPNPNPVFWSRDGATQWAVPGAPTGFIPHKGVFDPVNHVLIATNTGPGVDGSS 269
Db 64 F---APTRFVPHSTPLQATAYG-PACPOF-----NYPEELREITMAWNTPTPPAGES 114
QY 270 GDWKFESVTSGWTTRISPVSTDTANDYFGYSGLTIDRQHNTIMVATQISWMPDIIIFR 329
Db 115 EDCLNLNIY-----VPGTENTN-----KAVWV-----W-----IY- 139
QY 330 STDGGAT--WT--RIWDWTSYPNR-----SLRYVLDISAEPWLTFGVQPNPPVSPK 378
Db 140 ---GGALEYGWSFHLVDGASFAANDVIATVINTYNTNI-----LGFPAPAPLPITQRNL 191
```


QY	379	GWDEMAID-----PFNSDRMLYCT	399
		: :	:
Db	192	GFLDQRFALDWORNTAAAGGDPKRYTIFQSGAGRSVDVLLTSMHPNPPFRAAINESGV	251
QY	400	GATLYATNDLTK-WDSGGQIHIAPMVKGLEETAUNDLIS--PPSGAPLISALGDLG-QF	454
Db	252	ANYNPKGDLSEFWNT-----TVOALNCTTISIDILSCMRVYDIATLMNTIEQLGLGF	303
		: :	: :
QY	455	THA--DVTAVPTSTFTSPVTTCTSYDYAELNPSTIIVRAGSDPSPQPNDRHVAFTDGG	512
		: :	: :
Db	304	EYTLDNVTAV--YRSE--TARTTGDIARV-PVLV-----	332
QY	513	KNWFQSGEPGGVTGTGTVAAASADGSRFVAPGD-----PCQPVYAVGFGNSWAAS	563
		: :	: :
Db	333	-----GTVA--NDGLLFLVIGENDTQAYLEAIPNQPDLYQTLGLGAYPGS	375
QY	564	QGV--PAN-----AQIRSDRVNPKTFVLSNGTF-----YR	592
		:	:
Db	376	PGIGSPQDIAAIETEVRFQCPSAIVQADSRNRGIESWRY--YYNATFENLELFPGSVTH	434
QY	593	STDGCVTF--QPVAAG-----LPSSGAVGVMFHAV	620
		:	:
Db	435	SSEVGWFGTYPVASATALEAQTSKYMQGAWAAFAKNPMNGPGWKQVPNVAALG-----S	489
QY	621	PKRE-----GDLWLAASGLYHSTNGGSSWSAITGVSSAVNVGF	659
		:	:
Db	490	PGKAIQVDVSPATIDORCALYTRYTELGTIAPRTFGGSGGGSGKDNADVVVVGA	549
QY	660	GKS-----APGSSYPAVFVVGTGTGTVGAYS-----DDCGTWWLLINDDQ---	700
		:	:
Db	550	GLSGLETARKVQAAGLSCLVLEAMDRVGGKTLVQSGPGGTTINDILGAHW--INDNSQSE	607
QY	701	-----HOYGNNGQAITGDHANLRVYIGTNGRGIVYGD	733
Db	608	VSRLEFRFHLEGEI-RTTGN--SIHQAOGGTTTAA-PYGD	644
		:	:

```

RESULT 13
US-09-352-168-25
; Sequence 25, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duwick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 1000
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-168-25

```

Query Match 3.5%; Score 140.5; DB 4; Length 1000;
Best Local Similarity 20.9%; Pred. No. 0.041;
Matches 159; Conservative 69; Mismatches 224; Indels 309; Gaps 41;

Db 16 GLSLASGAPT VKI ---DAGMVVGTTTVPCTTATVSEFLGVPE-----AASPTR 63

QY 210 FVGVDANNPNVFSRGGATWQAVPGAPGTGTFPHKGKGFDPVNHVLYATPSNTGGPYDGSS 269

Db 64 F--APPTREVPWSTPLQATAYG-PACPOQF-----NYPEELREITMAWFNTPPPSAGES 114

QY 270 GDVWKFSVTSGTWTRISVPSTDTANDYFGYSLGTTIDRQHPNTINWATOISWNPDTIIFR 329

Db 115 EDCUNLNIY-----VEGTENTN-----KAVMV-----W-----Y- 139

QY 330 STDGGAT--WT--RIWDTYSYR-----SLRYVLDISAEPWLTFFGQPNPPVPSPKL 378

Db 140 --GGALEYGNSEFHYLDGASFAANQDVATVINYTN-----LCGPAAPQLPITQRL 191

QY 379 GWDEAMAID-----PFNSRMLXYGT 399

Db 192 GFLDQREALDVQNRNIAAAGDPRKVTIFQSGAGRSVDVLLTSMHPNPPFRAAIMESGV 251

QY 400 GATLYATNDLTK-WDSGGQIHAPWKVGLLEETAVNDLIS---PPSGAPLISALGDLG-CF 454

Db 252 ANYNFPKGDISEPMT-----TVOALNCTTSIDILSCMRVLDIATLMNTIEQLGLGF 303

QY 455 THA--DVTAVPSTIFTSPVFTTGSVDYAEPLNPSIIVRAGSPDPSPQNDRHVAFSTDGG 512

Db 304 EYILDNTAV---YRSE--TARTTGDIAV-PLIV----- 332

QY 513 KNWFGSGEPGVTGTGTTVAASADGSRFVWAPGD-----PGQPVVYAVGFGNSWAAS 563

Db 333 -----GTVA--NDGLLFVLGENDTQAYLEEAIPNQDLYOTLLGAYPIGS 375

QY 564 QGV--PAN-----AQRSDRVNPKTFYALSNGTF-----YR 592

Db 376 PGIGSPQDQIAAIEETEVRFOCPAIVAOBSRNRGIPSWRY-YNATFENLELFPQSEVYH 434

QY 593 STDGCVTF--OPVAAG-----LPSSGAVGVMFHAV 620

Db 435 SSEVGWFGTYPVASATALEAQTSKYMOGNAAFAKNPMPGCKQVNPVAALG-----S 489

QY 621 PGKE-----GDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGF 659

Db 490 PGKAIQVDVSPATIDQRCALYTRYVTELTGTIAPRTFGGGGGGGGSKDNADVAVVVGGA 549

QY 660 GKS-----APGSSYPAVFVVGTIGGTVGAYS-----DDCGTTWVLINDQ--- 700

Db 550 GLSGLETARKVQAAGLSCLVLEAMD RVGGKTLVSQSGPQRTTINDLGAAM--INDNSQSE 607

QY 701 -----HOQVGNWQAITGDHANLRVYITNGRGIVYGD 733

Db 508 VSRLEFRFHLEGEL-QRTTGN--SIHQADGTTTITA-PYGD 644

RESULT 14

US-08-570-311-14

; Sequence 14, Application US/08570311

; Patent No. 5824791.

; GENERAL INFORMATION:

; APPLICANT: Proguliske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/570,311
/ FILING DATE:
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA: US 08/353,485
/ FILING DATE: 09-DEC-1994
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/647,119
/ FILING DATE: 25-JAN-1991
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/241,640
/ FILING DATE: 08-SEP-1988
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Whitlock, Ted W.
/ REGISTRATION NUMBER: 36,965
/ REFERENCE/DOCKET NUMBER: UF15.C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (904) 375-8100
/ TELEFAX: (904) 372-5800
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2628 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-570-311-14

Query Match 3.4%; Score 136; DB 2; Length 2628;
Best Local Similarity 20.1%; Pred. No. 0.33;
Matches 178; Conservative 68; Mismatches 265; Indels 374; Gaps 52;

QY 69 VSTAADPINTNKV---WAAVAG-YTNSWD-----PNDGAILRSSD----- 104
DB 917 VNTVDVQVNPQNLTGSAVGKVTWKWDAPNGTNPNNPNNPGTTTISEFENGPAS 976
QY 105 -----QGATWQITPLPKLGGNPMGRGRLAV-----DPNNNDILY-- 142
DB 977 WKTIDADGCGNNWTTTPPP---GGT---SPAGHNSAICASSASYINFEQPNPDNVLVTP 1030
QY 143 -FGAPSGKGL--W-----RSTDGATWSQMTNPPD-----VGTIYIAP-- 177
DB 1031 ELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGDNFANALLEEVLTAFTVVTAPEA 1090
QY 178 -----TDTTGYQSDIQ---GVVWVAFD-----KSSSLGQASKTIFVG 212
DB 1091 IRGTRVQGTWYQTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGRADFTET-FES 1149
QY 213 VADPNPNPVFVS-----RDG-GATWQAVPCAPTGFTHPKG-----VDPVNVHL- 254
DB 1150 STHGEAPAEWTTIDADGCGWELCSGQLGWLTAHGGTNVVASFSWNGMALNPDNLYLS 1209
QY 255 -----YIATSNVTGGPYDSSGDVMKFSVTSCTWTRISVPSTDTANDY-----F 298
DB 1210 KVTGATKVKIYAVNDGPPGDHYAVMISKTGNTAGDFT-----VFEEETPNKNGGARF 1265
QY 299 GYSGLTIDRQHPNTIMVATQISWPDITIFRSTDGATWTRIMDWTSYNRSRLRYVL--D 356
DB 1266 GLS-TEADGAKPOSV-----W-----IERTVDLFA-GTKYVAFRHYNCSDNLNILLDD 1311
QY 357 ISAEPLWLTFCVQNPVPVPSKPLGWMDEAMDAIDPFNSDRMLYGTGATLYATNDLTKWDSGG 416
DB 1312 IQ-----FTMGGSPTP-----TDYTYTVVRDGTK----- 1335
QY 417 QIHIAPVWKGLETAVNLDLSPSPGAPLISALGDLGFTHADVTAVPSIFTSVPTTGT 476

DB 1336 -----IKEGLTETTFEE-----DGV-----ATGN-----HEYCVEVKYTAGVSPKECVNV 1375
QY 477 SVDYAELNP-----SIIVR-----AGSFDPSQPNDRHVAFS----- 508
DB 1376 TVDPVOFNPNQNLTGSAVGKVTWKWDAPNGTNPNNPNNPGTTTISEFENGPASWKT 1435
QY 509 --TDG-GKNWFQSGEPGGVT-TGGTVAASADGSRFVWAPCDPCQP----- 549
DB 1436 IDADGCGNNWTTTPPGGTSFAGHNSAICASSASYINFEQ-PQNPNDNLYLWTPPELSLPLNGG 1494
QY 550 -----VYVAVFGNS-----WASQGVPAANAQIRSDRVNP 579
DB 1495 TLTFWVCAQDANYASEHYAVYASSTGDNFANALLEEVLTAFTVVTAPETARGTRVQ- 1553
QY 580 KTFY-----ALSNGTFY-----RSTDGCVTFQPVAAQ----- 606
DB 1554 GTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGRADFTETFSSTHGEAPA 1613
QY 607 ----LPSGAGVGMFHAVPGKEGDLWLAASSGL--YHSTNGGS-----SWSA----- 647
DB 1614 EWTTIDADG-----DQGWELCSGQLGWLTAHGGTNVVASFSWNGMALNPDN 1661
QY 648 -----ITGVSS-----AVNVGKGSAPGSSYPVAFV--VGTIGG 679
DB 1662 YLISKDVTGATKVKIYAVNDGF---PGDHY-AVMISKGTGTNAG 1701

RESULT 15

US-09-352-159-29
; Sequence 29, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Gulliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER FILING DATE: 1998-07-25
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 1205
; TYPE: PRT
; ORGANISM: Unknown
US-09-352-159-29

Query Match 3.4%; Score 135.5; DB 4; Length 1205;
Best Local Similarity 21.3%; Pred. No. 0.12;
Matches 176; Conservative 71; Mismatches 230; Indels 349; Gaps 49;

QY 94 PNDGAILRSSD-----QGATWQITPLPKLGGNPMGRGRLAVDPNNNDILYFGAPS 147
DB 187 PQIDKYLKSSKIYAWPLOG--WQAT---FG-GGDHP-----PKSDLV-----PR 224
QY 148 GKGLMNST---DSGATWSQMTNPPDVGTIYIANPTDTGYQSDIQGVVWVAFDKSSSLGQ 204
DB 225 GSPEFAPTQKIDAGMV-----VGTUUTVP-GTTATVSEFLGVFFA----- 263
QY 205 ASKTFVGVGADPNPNPVFWSRDGATWQAVPGAPTGFTHPKGVDPVNVHLYTATSNITGCP 264
DB 264 ASPTRE---APTRPVWSTPLQATAYG-PACFQQP-----NYPEELREITMAWFNTPPP 314
QY 265 YDGSDDVMKFSVTSCTWTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQISWPD 324
DB 315 SAGESEDLNLIY-----VPGTENTN-----KAVMV-----W-- 342


```

QY 325 TIIFRSTDCGAT---WT---RWDWTSPNR-----SLRYVLDISAEPMWTFGVQVQPPV 373
Db 343 --IY-----GGALEYGWSFHYDGASFAANQDVIAVTINRTNI-----LGFPAAFPOLPI 391
QY 374 PPKLGMDEMAID-----PENS DR 394
Db 392 TORNLFGLDQRFALDWQORNIAAFQGGPRKVTIFGQSAGRSVDVLLTSMPHNPPFRAAI 451
QY 395 MLYGTGATLYATNDLTK-WDSGGQIHIAPMVKGLETAVNNDLIS---PPSGAPLISALGD 450
Db 452 MESGVANYNFPKGDLEPANT-----TVQALNCTTSIDILSCMRVRDLATIMNTIEQ 503
QY 451 LG-GFTHA--DVTAVPSTIFTSVDYAELENPSIIVRAGSFDPSOPNDRHVA 507
Db 504 LGLGFYETLDNVTAV---YRSE--TARTGDIARV-PVLV-----537
QY 508 STDGKNWFQSGPEPGVTGGTVAASADGSRFVWAPGD-----PGQPVVYAYVFGN 558
Db 538 -----GTVA--NDGLLFVLGENDTQAYLEEAIPNOPDLYQTLGA 575
QY 559 SWAASQGV--PAN-----AQIRSDRVNPKTFYALSNGTF-----590
Db 576 YPIGSPGIGSPQDQIAAIEVEVRFQCPQSAIVAODSRNRGIPSWRY-YYNATFENLELFP 634
QY 591 ---YRSTDCGVTF--QPVAAG-----LPSSGAVGV 615
Db 635 SEVYHSSEVGMVEGTYTPVASATALEAQTSKYMQCAWAFAKPNMNGPGWKQVNPVVAALG- 693
QY 616 MFHAVPGKE-----GDLWLAASSGLYHSTNGSSWSAITGVSSA 654
Db 694 -----SPKAIQVDVSPATIDQRCALYTRYVTELTGTIAPRTFGGGGGGGGSKDNVADV 749
QY 655 VNVGFGKS-----APGSSYPVAVVVTIGVTCAYRS-----DDCGTTWVNLND 698
Db 750 VVVGAGLSLETARKVOAAGSLCVLEAMDRVGGKTLVQSGPGRTTINDLGAW--IND 807
QY 699 DQ-----HOYGNMGQAITGDHANLRVYIGTNGRGIVYGD 733
Db 808 SNQSEVSRUFERFHELGEL-QRTTGN--SIHQADGCTTTA-PYGD 849

```

Search completed: July 2, 2002, 09:12:54
Job time: 62 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:13:57 ; Search time 54.67 Seconds
(without alignments)
1300.642 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTOPYTWSNVAIGGGFVD.....YICTNGRCIVYDGGAPSG 740
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97013	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical prote
6	202.5	5.0	2174	2 E92965	hypothetical glyci
7	189	4.7	3972	2 S73251	hypothetical prote
8	186.5	4.6	4199	2 S78412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RFX faml
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 AB3528	extracellular seri
17	166.5	4.1	902	2 H87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1884	2 S10789	amylose A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD0835	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

30	157	3.9	809	2 A55547	quinate-shikimate
31	156.5	3.9	2204	2 A70524	probable PPE prote
32	156	3.9	3157	2 B70969	probable PPE prote
33	155	3.8	4180	2 G83559	hypothetical prote
34	154.5	3.8	1032	2 T34433	hypothetical prote
35	154.5	3.8	3570	2 T45025	mucin MUC5B, trach
36	153	3.8	1055	2 A87364	OmpA-related prote
37	133	3.8	1821	2 AG2335	hypothetical prote
38	153	3.8	13288	2 T03099	mucin, submaxillar
39	152.5	3.8	3016	2 S77300	hypothetical prote
40	152	3.8	348	2 T35248	probable oxidoredu
41	152	3.8	1049	2 T42045	beta transducin-li
42	152	3.8	1196	2 A29130	beta-amylase (EC 3
43	152	3.8	3472	2 T31308	hypothetical prote
44	151.5	3.8	820	2 B72575	hypothetical prote
45	151.5	3.8	1145	2 B75625	hypothetical prote

ALIGNMENTS

RESULT 1
T35237
probable secreted cellulase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35237
R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21572
A;Accession: T35237
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-890 <SEE>
A;Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GMO00070; SCOEDB:SC5C7.30C
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC5C7.30C

Query Match	58.7%;	Score	2371;	DB 2;	Length	890;			
Best Local Similarity	57.7%;	Pred. No.	1.3e-131;						
Matches	431;	Conservative	109;	Mismatches	183;	Indels	24;	Gaps	10;
QY	6	YTSNVAIGGGFVDGIVFNEGAPGILYVTDIGMVRWDAANGRWITPLLDWGVNNWGY	65						
Db	41	YTKNARIDGGFVPGIVFNRTEKDLAYARTDIGGAYRWQESHTWPTLLDHVGDWDMGH	100						
QY	66	NGVYSTAADPINTNKYAAVGMKTSNDNDGAILRSSDQATWQITPLPKLGGNNMGR	125						
Db	101	TGVVALASDAVDPRVAAVGTNTNDMDPTNGAVLRSDRGASWEKADLPKLGNNMGR	160						
QY	126	GMGERLAVDPNNNIIYFGAPSGKGLWRSTDGATNSQMTNFPDVGTYIANPTDITTCQS	185						
Db	161	GMGERLAVDPHDNDVLYLGLAPSGHGLWRSTDAGVTNSEVTAFFPNPNYAQDPNDTSGYAS	220						
QY	186	DIQGVVWVAFDKSS-SSLGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHK	244						
Db	221	DNQGITWVTFDESTGGGAGTATRLVYGVADKENAVYRSTDAGATWERLAGQPTGYLAHK	280						
QY	245	GVFDPVNHVLIATSTNGGPGYDGSDDVWKFVSTSGTWTRISVPVSTDTANDFGYSGLT	304						
Db	281	GVLDANGYLIAYSDTGGPGYDGGKRLRYATATGTWTDISPAEAADT---YFGSGLT	337						
QY	305	IDRQHPNTIMVATQISWWPDTIIFRSTDGGATWTRIDWTSYPNRSRLRYVLDISAEPLWT	364						
Db	338	VDQRQCTVMATAYSSWWPDTQIFRSTDGATWQAVPGAPTGFIPHK	397						
QY	365	FGVQPNPPVPSKLGHWDEAMADPPNSDRMLYGTGATLYATNDLTKW-DSGGQIHTIAPM	423						
Db	398	WGANPAPPQPTPKLGWMTSALEDIPFDSRMYGTGATVYGTENLTWNDDDEGGTFAVEPM	457						
QY	424	VKLEETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGTSDYAE	483						

Db 458 VRLEETAVNDLASPPSGAPLLSALGVDGGRHTSLTEVPSMMYTSNFSTTSLDFAET 517
 QY 484 NPSIIVRAGSDFDSSQPNDRHVAFASTDDGKNWFQGGSPGGVTTGGTVAASADGSRFVWAP 543
 Db 518 KPDVVVRAGNLD--SGP--HIAFSTDNGANFGGTDPSGVSAGGVAAGADGSRFVWSP 572
 QY 544 GDFQOPVYVAVGFGNSWAASQGVPAQAIRSDRVNPKFTYALNGTFFYRSTDDGGVTFQPV 603
 Db 573 --EGAGVOYTTGCTSQWASTGLPAGAVESDRVPATFYGFKSGRFYVSTDDGATFTAS 630
 QY 604 AA-GLPSSGAVGVNFAVPCKEGDLMLAASS-----GLYHSTNGGSSWSAITGVSSAVNV 657
 Db 631 AATGLPAGD--GYRFKALPGEGDVLWLAGAADGPGYGLWHSSTDDGGTFFRLPGVDAADTV 688
 QY 658 GFKSGAPSSYPVAVVVGTTGGVTVGAYRSDCGTTTWYLLINDDQHOYGNMGOAITGDHANL 717
 Db 689 GFKGAAPGASYQLFSAEIGGVGRGIFRSTDAAGATWRVNDADAHQWGTGAAITGDPVRY 748
 QY 718 RRVYIGTNGRGIVYGDV----GGAPSG 740
 Db 749 GRVYVATNGRGIVYGDVSDRGGCTDPG 775

RESULT 2
 D97013
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: D97013
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: D97013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:g15023820; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0919

Query Match 49.8%; Score 2009; DB 2; Length 839;
 Best Local Similarity 50.9%; Pred. No. 2.2e-110;
 Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;

QY 3 TQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRTDIGGMRYRWDAAANGRWIPLLDWVGWNN 62
 Db 37 SQGYKWDNAKIGAGGYVPAVIFNKTEKDLIYARTDMGGAYRWDKANNKWIPITD--GFSD 94
 QY 63 WGVNGVYSIAADPINTNKVAAVGMVNTSNDPNDGATLLRSSDQATWQITPLPLKGGNM 122
 Db 95 WTMJGESIATDPTDITNRVYIAGLTNDWQDENAYILLSSQDKNGTWKRYQLPFKVGNN 154
 QY 123 PGRMGRLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDITG 182
 Db 155 PGRNMGRLQIDPNDKILYLGARSGLWKSSEDYGTWSKVDNFPDPTDGYVDQPNQNE-- 212
 QY 183 YQSDIQGVVWVAFDKSSSSLGQASKITFYGVAD--PNNPFWRSRGATWQAVPGAPTGI 241
 Db 213 YTDKVGVMVETEDPSTGTGSPQTMYVGAADTKGNINVTNDGGKTSVAVKQPKGYL 272
 QY 242 PHKGVDPVNVHLYIATSNRTGGYDGSQGVWKFVSFVTSFSGTWTIRISPVSPSTDANDFGYS 301
 Db 273 PHHGIL--ASDGMILYISYNTCGYDGSQGVWKFVSFVTSFSGTWTIRISPVSPSTDANDFGYS 328
 QY 302 GLTIDRHPNTIMVATQISWPTIIFRSTDDGATWTRIMDWTSYNRSRYVLDISAEP 361
 Db 329 GISVDAQNPNNVVATLNRWPPDEEYRSTDACTWKPINDWNGYPNRTLYNLDISAQ 388

QY 362 WLTFG-VQPNPPVPSPKLGHMDEAMAIIDPNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420
 Db 389 WLDWGTGTGVTTPDPLVKLGMMGDLEIDPNSDRMFYGTGATLYGTDLTNWDKGNVDI 448
 QY 421 APVWGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTTGTGSDY 480
 Db 449 SVKANGIECAVNDVVVYPTKGAQLLSAVGDCGFIHDDITKVPKMMTTPNFSAITSDY 508
 QY 481 AELNPSIIVRAGSDFDSSQPNDRHVAFASTDDGKNWFQ--GSEPGGVTTGGTVAASADGSRF 539
 Db 509 AESVPNFVVRGVNDVTSKNQDKDCGISYDGGKNWFSAGSNIQVYKAGTVAAGADAKTI 568
 QY 540 VWAPGDPQGVVYVAVGFGNSWAASQGVPAQAIRSDRVNPKFTYALNGTFFYRSTDDGGV 599
 Db 569 VWSF-BEGANAAYSTONGNKWTPCSGLPQCAKVRSDRVNPKFTYALNGTFFYRSTDDAG 627
 QY 600 F-OPVAAAGLSSGAVGVNFAVPCKEGDLMLA--ASSGLYHSTNGGSSWSAITGVSSAVNV 657
 Db 628 FTOSSTGLPTKGR-GI-FKTVIGHEGDIWAGKGLWHSSTDSGATFTKVGVDASDVT 685
 QY 658 GFKSGAPSSYPVAVVVGTTGGVTVGAYRSDCGTTTWYLLINDDQHOYGNMGOAITGDHANL 717
 Db 686 GLCKSKTDDGYPALYMDATIDGTAGIFRSDEGATWVRINDADAHQYGSYDTCITGDPNKY 745
 QY 718 RRVYIGTNGRGIVYGDIGA 737
 Db 746 GRVYVATNGRGIVYGDIGS 765

RESULT 3
 T00349
 C:Species: Aspergillus aculeatus
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
 C:Accession: T00349
 R:Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
 submitted to the EMBL Data Library, June 1998
 A:Description: Avicelase III from Aspergillus aculeatus.
 A:Reference number: Z14141
 A:Accession: T00349
 A:Status: preliminary;
 A:Molecule type: mRNA
 A:Residues: 1-856 <ARA>
 A:Cross-references: EMBL:AB015511; NID:d1199887; PID:d1029971
 C:Genetics:
 A:Gene: avIII
 C:Superfamily: fungal cellulose-binding domain homology
 F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 41.6%; Score 1680; DB 2; Length 856;
 Best Local Similarity 46.2%; Pred. No. 4.8e-91;
 Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTQPTWSNVAI--GGGFFVDGIVFNEGAPGILYVTRTDIGGMRYRWDAAANGRWIPLLDWVG 59
 Db 21 AASQAYTWKNNVTVGGGGFTPGVFNPSAKGVAYARTDIGGAYRLN--SDTWTPLMDWVG 79
 QY 60 ---WNNWNGVYSIAADPINTNKVAAVGMVNTSNDPNDGATLLRSSDQATWQITPLPF 116
 Db 80 NDTWHDW---GIDALADPDVDTRDVRVAVGMVNTSNDPNDGATLLRSSDQATWQITPLPF 136
 QY 117 KLGNNPGRMGRLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 176
 Db 137 KVGNNPGRMGRLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 196
 QY 177 PTDTTGYSQDIQGVVWVAFDKSSSSLGQASKITFYGVAD--PNNPFWRSRGATWQAVPGA 236
 Db 197 SSST--YTSDPVGIWVTFDSTSGSSGSAATPRIFGVVADAGKSVFKSESDAGATWAWYSGE 254
 QY 237 PT-GFIPHKGVDPVNVHLYIATSNRTGGYDGSQGVWKFVSFVTSFSGTWTIRISPVSPSTDAN 295
 Db 255 PQYGFPHKGVLSPEKTLIISYANGAGPYDGTNGVHKYNTISGWWTDISP---TSLAS 311

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Accession: S75251
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S:Reference number: S74322; MUID:97061201
A:Accession: S75251
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3972 <KAN>
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAAL1765.1; PID:g165224
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr1028

Query Match 4.7%; Score 189; DB 2: Length 3972;
Best Local Similarity 19.3%; Pred. NO. 0.013;
Matches 166; Conservative 93; Mismatches 276; Indels 326; Gaps 42;

QY 13 IGGGGFVGVFNEGAPGILYVRTDIGMYRMDAANG-----RWIPLLDW- 57
Db 2458 VNGDGFADVISGGSPAGGVLIIFGNTKDLL--DAALGTDLIIISVENAQVKEFVALGDFD 2515
QY 58 -----VGNWNGYNG---VVSTAADPINTKNVAAVGMTNSWD-- 93
Db 2516 GDGLADFGVDDGNGFVLVLGSPQLSGSLVLDSTLPNLNFNQAGVGDGNGYDDF 2575
QY 94 ---PNDGAILRSSDGCATWQITPLPKLGNMP-----GRGMCERLAVDPN-- 136
Db 2576 VLOGPNSTIAYGNANGTLTDSPLTF--GNPFPLSSFGIDLNGNKEIVAGQPNLN 2633
QY 137 -----NDNIFYGAPSGKGLWRST-----DSG-ATWSQMTNFPDVGTYIANP 177
Db 2634 PVPNIGGFGGLGYFTYEAGNAVLPQTPNPNASVTEASGLSWGQI--SPNQYAAQAGVP 2692
QY 178 TDTT--GY-----QSDIQGVVW-----VAFDKSS----- 200
Db 2693 SPATLQGLWLYOAFYGINERISTKDSYIYIORSRGGVSWENLTQVPLDSNGTPIDLKNLP 2752
QY 201 -SLGQASKTIFVGVADPNPNPVFWSRDG-----GATWQAVP-----GAPT--GFTPHK 244
Db 2753 PSITAYNGTLYLFTADNGQV--WVAEGVNTNANSGLILNAVPIQNASNGPTLVAFNDEL 2811
QY 245 GVF-----DPNVHLYIATSNGTGPGYDGGSDVWKFVSCTGTRISPVPTDTPANDYFGY 300
Db 2812 YVFFVKDASNDILYSSSSNPG-----SSSG--WDGTSTVLTFSDVNOATN-----FPL 2858
QY 301 SGLTIDRQHPNTIMWATQISWMPDTIIFRSTDGATWTRI-----WD----- 342
Db 2859 SATVVPGLDGTILAVA-----FRSNSPATWVGLLNSDVTNWOGSALTEQVDA 2907
QY 343 -----WTSYPNRSRLRYL-----DISAEPWLATFGVQPNPVPSPKL 378
Db 2908 NSQVSLTVVDGTYLYLFTSTSEASVATSTDLGNWGDITLIPW-----DDGNL 2956
QY 379 GWDE-----AMADFPNSDRMLYGTGATLYATNDLTKWDSGGQTH----- 419
Db 2957 GGVAIFLNFQSFILSLNQSNNESLLFAFSNSLFEPNQASRW--GEQVRDIDGFDGDIAD 3014
QY 420 ---IAPVKGLETAVNDLISPPSGAPLISALGDL-----GGFTHADVTPVSTIFTSPVF 472
Db 3015 LAVIAPGRYRLLOPILDY-----PAINNLGVTIYIEESGISVNDPDPVLAAPDL 3067
QY 473 TTGTSVDVYAEINSLIVRAGSDP--SSQP-----NDRHVAFTDGGKNW-- 515
Db 3068 PQETIFELLEITPTGVNGDGFDDLLISAPLTPVIAQGFDPVNGDGVSVWVFGGTHWG 3127
QY 516 -FOGSEP-----GGVTG-----GTVAASADG--SRF-VW 541
Db 3128 EYTANSPFGLGNLANNQTNNSQNFYGFVTTGLPRSQAGISISGGADVNGDGFSDFALG 3187

QY 542 APGD-----PGQPVYAVGFGNSWAASQG---VPANAQTRSDR 576
Db 3188 APCNFNLNLYVLFSGDFTNQVNLGTIGDDVLMGLSGTGEIFVAGQGDQDIYTNNGVDIVY 3247
QY 577 VNPKTFYALNSNGTFYRSTGGS 597
Db 3248 AGPNDFFVVTDTNFRRLDGG 3268

RESULT 8
S76412
hypothetical protein.slr0408 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76412
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76412
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4199 <KAN>
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAAL8541.1; PID:g165 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr0408

Query Match 4.6%; Score 186.5; DB 2: Length 4199;
Best Local Similarity 19.2%; Pred. NO. 0.019;
Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

QY 9 SNVAIGGGFVGDIV-----FNEGAPGILYVRTDIGMYRMDAANGRWIPLL 55
Db 2261 ADVASDNGFVIDGNLGNPPTTFTTSTQYIDTTPAILNGSNLYLAKYKGGGNQ---IY 2317
QY 56 DWYGMNN--MGYNGVVSIAADPINTKNV--AAGVMYNSW-----DPNDGA-IIRSSDQGA 107
Db 2318 FTVSTNNGOSWSEVOL---PQSAQTIFPPAIAFFNNVLYLAYVDGNGNLNITSDQGG 2374
QY 108 TWQITPLPKLGNMPC-----RG-MGERLAVDPNNDNILEGAPSGKGLRSTDSGA 159
Db 2375 TWN---APLALGTSSTPPLFVYQGLTSLLLFAANNSTSTVLQFYLNSSNEWIYANEIGS 2431
QY 160 TWSQMT-----NFPDVGTYIANPTDTTGYQSDIQGVWVAF--- 195
Db 2432 NQTAISAISATVLGDTLYLYKGGTRNTPTSLDYIISTTN-----ADLSANDHSSPIPG 2486
QY 196 --DKSSSSLGQASKTIFVGVADPNPNPV-FWSRDGGATWQAVPGAPTGF---PHKGVFDP 249
Db 2487 VSSQGGPSLTNDGTNLVLSYLDSSNQLNFVSSGNGINWSS-POVITNNISQSPPAIAF-- 2543
QY 250 VNHVLIATSNCTGPGDGGSDGVWKFVSCTGTRISPVST-----DTAN 295
Db 2544 ANNELYS-----YPGQGS-QELNVTs-----FPLPFTGSILNGSLVRLFGDVNG 2589
QY 296 DYFG--YSGLTIDRQHPNTIM-----VATQISWMPDTIIFRSTDGGAWTRIDWTSY 346
Db 2590 DGFADYFSGGT-----NAGAIIFGNSTKDLTTTASGSEDLVI-----SV 2628
QY 347 PNRSLRYVL---DISAEPWLATFGV-----QPNPVPVPSPKL 378
Db 2629 PNATLRDVTSGVDFNGDGKIDGLDNGNFYVVLGNTSLGDLTKLSITSSSPVIVNOV 2688
QY 379 GWDEAMADFPNSDR-----MLYGT-----GATLYATNDLTKWD----- 413
Db 2689 GGVTKSAIGDYGDDGVLLMGDNGCNQVAGNSTGVLSNFTNIDYPTETQTTATGVDLN 2748
QY 414 -----SGGQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGTHAD 458


```
Db 2749 SDGIPEIAIGSDEKRIAGQISTSGSFLPTTSSVINTLAAANOLENIGDFNGDIAD 2808
QY 459 VTAVPSIFTS-----PVFTTGTSDVYAEIN-----PSLIIVRAGSFDPS 497
Db 2809 LAVLASWYAAAIAGEPNLNPLSRPGNOGGVFIFYGNSNGLSNTAQPDVILAAPTNPS 2868
QY 498 SQ-----PNDHRHAFSTDGKNW----- 515
Db 2869 GQISTVQLSRIAQAGDVNGDFDILLISSPYTDAENNOQGVFVFGGDDRNQPFDLGQ 2928
QY 516 -----FGSPEPGVTTGGTVAASA-----DGSRF-----VWAPG----- 544
Db 2929 LRANQSGSNPRFAIDGSPNSQAGIALNGGDIINGDFADFIIGAPENNLQYNQOIVF 2988
QY 545 -----DPGQPVYAVG-----FGNSWAASQGVPAQAQIRSDRYN-----PKTFYAL 595
Db 2989 IENGESDDDKYSILYLDGNOTIOMGGDWQANO-VWNT-QVATWNNSRPEPAVIGQ 3046
QY 586 SNGTFYRSTDGGVTFQ-----PVAAGLPSSGAVGVMFHFAVPGK 623
Db 3047 SNGDIWYPCGNQWQSWGKLPAEINELAVNNTSGNPQIIAGLGKG--GIEYY----- 3099
QY 624 EGDWLAAASGLYHSTNGSSW-SAITGVSSAVNVGFGKSAPGSSYPVAVFVGTIGVVG 682
Db 3100 NGSTWV--NNGPYQ---GDGWSRAITQMA---VOMGED--GS--PSQIVVGLADCAVI 3145
QY 683 AYRSDDCGTTWLINDDHOQYGNWGOAITG-----DHANLARVYIGTN 725
Db 3146 YYNTQ---SCWRTIN-----NFGSKVTSLSVOMQESAPNIVVGLDNSEV-QYOGSN 3194
QY 726 GRGIVYGDIG 735
Db 3195 GVWTQFHDG 3204

RESULT 9
T13256
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T
C:Species: Lactococcus lactis phage BK5-T
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C:Accession: T13256
R:Boyce, J.D.; Davidson, B.E.; Hillier, A.J.
Appl. Environ. Microbiol. 61, 4089-4098, 1995
A:Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and
A:Reference number: Z17646; MUID:96064422
A:Accession: T13256
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1904 <BOY>
A:Cross-references: EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579.1

Query Match 4.5%; Score 182; DB 2; Length 1904;
Best Local Similarity 20.7%; Pred. No. 0.012;
Matches 189; Conservative 83; Mismatches 260; Indels 380; Gaps 55;

QY 1 ATTQPYTWSNVAIGGGFVDFIVFNEGAPG-----ILYVRTDIGMYRWDANGR 50
Db 838 AKPSDYSWS--LIRGNDGKDGATGKDGACGVGKIKTTVITVALSSG---TDKPNTG 891
QY 51 W--IPLLDWGNWNGYGVVSIADPINTNKVAAVGYMYSWDPNDGAILRSSDQGA 107
Db 892 WTSQVPTL-----VKGOYLTKTWVT---YTDG-----SSETG- 921
QY 108 TWOTLTPFLKLGMPGMRGERLAVDPNNDNITLYFCAPSGKGLWRSTDGATW-SQMTN 166
Db 922 -YSVTYI--AKDGNNGNDIAGKDGKGVKIKTTTYAVGTSG-----TTAPASGNVSOVPN 973
QY 167 FP-----DVGVTIANPTDTTGIQYSDIQGVVWVAFDKSSSLGQASKTIFVGVADPN 218
Db 974 VPAGQFLWTKTVWYTDN-TSETGYSVAMVMGVKDGKDPGNNGTN-----GIA---- 1020
QY 219 PVFWSRDG-----GATWQAVPG----APTG----FIP--HKGVF-----DP 249
```

```
Db 1021 ----GKDGKIGKATAIYQASPNGTTAPTGTWSASVPPVAKGSFLWTRTITWYTDNTTET 1076
QY 250 VNIHLVIAIANTGCGPYDGGSGDVWKFESVTSCTWTRISLPSPSTANDYFGYGLTIDRQH 309
Db 1077 GYAVAYMGTNGNG--HDGFPG-----KDGTKIKTTIT-----YAGTSGTTP 1118
QY 310 PNTIMVATQISWPPDIIIFRSTDGGATWTR-IWDMTSYPNRSLRYVLDISAEPWLTFGVQ 368
Db 1119 PN-----NGMTSTVP-TVAEGNLTWTKTVWYTD--NTS----- 1149
QY 369 PNPPVPSPKLGWDEAMAI-----DPFNSDRMLYGTGATLYATNDLTKWDGGQIHAPM 423
Db 1150 -----ETGYSVAMVMGVKDGKDPGN--GNGIAGKDG----- 1180
QY 424 VKGLETAVNDLISP-PSGAPLISALDGLGGFTHADYAVP--STIFTSPVFT----- 473
Db 1181 -KGIKATAIYQASPNGTTAPT-----GTWSASVPPVAKGSFLWTRITWYTDNTTE 1231
QY 474 TGTSDVYAEIN-----PSIIVRAGSEDPSSQPNDRHFAVSTDGKKNWF 516
Db 1232 TGAVAYMGTNGNGNHDGFPKDGTKIKTTITYAGTSGTTPPN--GW- 1279
QY 517 QGSEPGGVTTGGTVAASADG---SRFVW-----APGDPGPVYVAVG 555
Db 1280 -----TSTVPTVAEGNLTWTKTVWYTDNTSETGYSVAMVMGVKDGKDP----- 1323
QY 556 FGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFYRSTDGGVTFQFVAAG----- 606
Db 1324 -GNNI--GTNGI-AGKDGKIGKATAIYQASPNCTTAPTGTWSASVPPVAKGSFLWTRTIW 1379
QY 607 -----LPSSGAVGVN-----FHAVPGKEG-----LWLAASSGLYHSTNGSS-- 644
Db 1380 TYTDNTETGYAVAYMGTNGNNGHDGFPKDGTKIKTTITYAGTSGTTPPNNGTSTV 1439
QY 645 -----W-----SAITGVSSA-VNVGFKSAPGSSYPVAVFVGTIGGVGTG-- 682
Db 1440 PTVAEGNLTWTKTVWYTDNTSETGYSVAMVMGVKDGKDPGN--GT-NGIAGKDG 1491
QY 683 -----AYRSDDCGTTWLINDDHOQYGNWGOAI---TGDHANLRV----- 720
Db 1492 GKGIKATAIYQASPNGTT-----APTGTWSASVPPVAKGSFLWTRITWYTDNTTE 1543
QY 721 -----YIGTNG 726
Db 1544 TGAVAYMGTNG 1555

RESULT 10
F70846
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70846
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: F70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2523 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17115.1; PID:e125
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 4.3%; Score 175.5; DB 2; Length 2523;
Best Local Similarity 21.8%; Pred. No. 0.044;
```


Db 4567 IQVTGTLTWYGSII-----VAFDKEGDYLDWFWDAGGPDST 4605
QY 523 -----GVTTGGTVAASADGSRFVWAPDGPQPVYVAVGEGNSWAASQGVPAANAQIRSDR 576
Db 4606 FLNNAGTLVGNSTTSNGSGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4655
QY 577 V-----NPKTFYALS-----NGTFYRSTGGVTFFQVAAAGLPSGAGVYMFHAPVCK-EG 625
Db 4656 VQHTYLNLYTLSSLLINGNGTFVWGQNTNTFSLGAGSGAMSSSVMTWADFDCGDGM 4715
QY 626 DLWLAASSG-----LYHSTNG----- 641
Db 4716 DLFLPASQGRANYGSLLENTNGVLGCPVAVGATATTYASQFSLAVDWNHDLMDIARIAQ 4775
QY 642 -GSSWAIATGVSSAVNVGKSPAGSSYPAPVFFVGGIGVTT-----GAYR-----SDDCG 690
Db 4776 TQGSY-LYTNVSNASN--WTQSAALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4827
QY 691 TTWVLINDQHOYGNNGQAITGDHANLRVYIGTNGRGVIYGD 733
Db 4828 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4860

RESULT 14
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033965.1; PID:g13360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs0542

Query Match 4.2%; Score 168.5; DB 2; Length 5291;
Best Local Similarity 20.1%; Pred. No. 0.3;
Matches 190; Conservative 82; Mismatches 270; Indels 401; Gaps 46;

QY 27 GAPCILYVVRTDIGMYRWDAANGRWIPLLDWVGNWNGVYVSIADPINTNKVWAAVG 86
Db 4186 GANEFAQISTDNGA-----TWVN-----VTVAADSLN----- 4212

QY 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGGNMPGRGMGERLAVDPNN----- 137
Db 4213 -----WSYVDGRLTNGT--TTWQVRV--DLAGNV-GATSSOSALIDTVPNAQVLTIAS 4262
QY 138 -----DNILYF-----GAPSGKLMRSTDSGATSQMTNFPDVGVIYAN 176
Db 4263 ISDTGSSATDFITSDTMTLTGSLGAGLASGEVAQISLDSGATWTLTATNGTQWYTD 4322
QY 177 PTDTTG--YQSDIQGVWVAFDKSSSLGQASKTIFVGVADPN---NPVFWSRDGGATW 230
Db 4323 RTLTGDSYVYQVRV-----LDLAGNTGPVSKTVVVDVTPINPTAPTIVSYTDDVGQRQ 4375

QY 231 QAVPGAPTGFIPKHGVDPNVHLYIATSNTPGPGYDG-----SSGDV----- 272
Db 4376 GTLSSSQ-----ATDDTTPLLNGVLSPALASGEVYLYRNGLLGA 4416
QY 273 -----WKFS---VTSGWT---RISPVSTDDTANDYFGYSLGIDTIDRHPNTIMVAT- 317
Db 4417 VTMVGNLNTWYSDSLVSGAYTYSARVVDLAGNITSSDF---VLAVDTSTPTLAQITS 4473
QY 318 -----QISHW-----PDT--- 325

Db 4474 QITRDTPIISGIVTAALASGOYVEVWINGKYITSEPGAVVVVDPAHNTWYVQLPDTDAL 4533
QY 326 -----IIFRSTDGGATWTRIWDTWTSYPNRSRYVL 355
Db 4534 TVSATAYTVAQVKSSAGNANNISNGTVTVNAAIDYPTPTWTTASKTAW---GLTYGL 4590
QY 356 DISAEPLWTFGQPNPVPSPKLGWMDDEAMIDPFNSDRMLYGTGATLVATNDLTWKDSG 415
Db 4591 D-SHGMMTVLANQQVMQSTDP-LTWSKTALT-----LYQSGNN-YATSIADIYDRN 4638
QY 416 GOIHAPMVKGLEETA VNDLISPPSGAPLIS-----ALGDLGGFT-HADVTAVPSTIFTSP 470
Db 4639 G-----TGDLFTTRDDYCTGYINGFTNNGDGT-----FSSA 4669
QY 471 V-FTTGTSDYDAELNPSIIVRAGSFPDPSPQNDNRHVFST--DGGKNWFOGSEPG----- 522
Db 4670 IQVTGTLTWYGSII-----VAFDKEGDYLDWFWDAGGPDST 4708
QY 523 -----GVTTGGTVAASADGSRFVWAPDGPQPVYVAVGEGNSWAASQGVPAANAQIRSDR 576
Db 4709 FLNNAGTLVGNSTTSNGSGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4758
QY 577 V-----NPKTFYALS-----NGTFYRSTGGVTFFQVAAAGLPSGAGVYMFHAPVCK-EG 625
Db 4759 VQHTYLNLYTLSSLLINGNGTFVWGQNTNTFSLGAGSGAMSSSVMTWADFDCGDGM 4818
QY 626 DLWLAASSG-----LYHSTNG----- 641
Db 4819 DLFLPASQGRANYGSLLENTNGVLGCPVAVGATATTYASQFSLAVDWNHDLMDIARIAQ 4878
QY 642 -GSSWAIATGVSSAVNVGKSPAGSSYPAPVFFVGGIGVTT-----GAYR-----SDDCG 690
Db 4879 TQGSY-LYTNVSNASN--WTQSAALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4930
QY 691 TTWVLINDQHOYGNNGQAITGDHANLRVYIGTNGRGVIYGD 733
Db 4931 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4963

RESULT 15
A28452
proteoglycan core protein precursor, cartilage - rat
N:Alternate names: aggrecan
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 21-Jan-2000
C:Accession: A92623; A28453; A28454; A28455
R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 262, 17757-17767, 1987
A:Title: Complete primary structure of the rat cartilage proteoglycan core protein de
A:Reference number: A92623; MUID:88087070
A:Accession: A92623
A:Molecule type: mRNA
A:Residues: 1-2124 <DOE>
R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 263, 10040a, 1988
A:Reference number: A30069
A:Contents: annotation; revision to residue 698
R:Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 261, 8108-8111, 1986
A:Title: Partial cDNA sequence encoding a globular domain at the C terminus of the ra
A:Reference number: A23835; MUID:86250698
A:Accession: A23835
A:Molecule type: mRNA
A:Residues: 1856-2124 <DO2>
R:Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 262, 17768-17778, 1987
A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-t
A:Reference number: A28453; MUID:88087071
A:Accession: A28453
A:Molecule type: protein
A:Residues: 20-37, 'W', 39-60, 'E', 62-64, 'X', 66-69; 70-83; 84, 89-148, 'L', 150-238, 'S', 240, 'E'

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:53 ; Search time 28.05 Seconds
(without alignments)
1021.478 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPTWNSVAIGGGGFVD.....YICNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	168	4.2	2124	1 PGCA_RAT	P07897 rattus norv
2	167.5	4.2	1119	1 ALS3_CANAL	O74623 candida alb
3	166	4.1	1468	1 N153_RAT	P49791 rattus norv
4	157	3.9	809	1 QUITA_ACICA	Q59086 acinetobact
5	153	3.8	1150	1 APWU_PIG	P12021 sus scrofa
6	152	3.8	1196	1 AMTB_PAEPO	P21543 paenibacill
7	149	3.7	2132	1 PGCA_MOUSE	Q61282 mus musculu
8	148.5	3.7	776	1 ISOA_PSEAY	P10342 pseudomonas
9	148.5	3.7	776	1 ISOA_PSESP	P26501 pseudomonas
10	148.5	3.7	1953	1 BIGA_SALTY	P25927 salmonella
11	146	3.6	2109	1 PGCA_CHICK	P07898 gallus gall
12	145.5	3.6	790	1 QUITA_XANCU	Q9xg78 xanthomonas
13	145	3.6	342	1 Y48L_SVNY3	P73069 synechocyst
14	145	3.6	747	1 GUND_CELFI	P50400 cellulomona
15	143.5	3.6	937	1 NU98_RAT	P49793 rattus norv
16	142.5	3.5	995	1 AGAA_VIBS7	P48839 vibrio sp.
17	142	3.5	647	1 NANH_MICVI	Q02834 micromonosp
18	142	3.5	757	1 DHET_GLUOX	O05542 gluconobact
19	142	3.5	760	1 YBIL_ECOLI	P75780 escherichia
20	139.5	3.5	1045	1 GUNB_CELFI	P26225 cellulomona
21	139.5	3.5	2333	1 PGCA_CANFA	Q28343 canis famli
22	139	3.4	3063	1 CAIC_HUMAN	Q99715 homo sapien
23	138	3.4	827	1 XANP_XANS2	Q60106 xanthomonas
24	137.5	3.4	872	1 GUXA_CELFI	P50401 cellulomona
25	137.5	3.4	1103	1 VG37_BPARI	Q990b5 bacterioph
26	137.5	3.4	2415	1 PGCA_HUMAN	P16112 homo sapien
27	136.5	3.4	522	1 NU62_HUMAN	P37198 homo sapien
28	136.5	3.4	781	1 NANH_VIBCH	P37060 vibrio chol
29	136	3.4	2628	1 HAGA_PORGI	Q51845 porphyromon
30	134.5	3.3	2344	1 POLN_RHDV	P27410 rabbit hemo
31	134	3.3	890	1 BCN5_CLOPE	P08696 clostridium
32	134	3.3	1014	1 NANH_CLOSE	P29767 clostridium
33	133.5	3.3	966	1 FIB1_PETMA	P02674 petromyzon

ALIGNMENTS

RESULT 1

ID	PGCA_RAT	STANDARD;	PRT;	2124 AA.
AC	P07897;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).			
DE	protein) (CSPCP).			
GN	AGC1 OR AGC.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88087070; PubMed=3693370;			
RA	Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;			
RT	"Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones."			
RL	J. Biol. Chem. 262:17757-17767(1987).			
RN	[2]			
RP	REVISION TO 698.			
RA	Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;			
RN	J. Biol. Chem. 263:10040-10040(1988).			
RL	[3]			
RP	SEQUENCE OF 1856-2124 FROM N.A.			
RX	MEDLINE=86250898; PubMed=2424893;			
RA	Doegge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;			
RT	"Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan."			
RL	J. Biol. Chem. 261:8108-8111(1986).			
CC	-!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.			
CC	-!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).			
CC	-!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.			
CC	-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.			
CC	-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 4 LINK DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.			
CC	-!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -			

34	133	3.3	920	1 NU98_HUMAN	P52948 homo sapien
35	132	3.3	886	1 VGP3_EBVA8	Q07284 epstein-bar
36	132	3.3	1433	1 SUBF_BACSU	P16397 bacillus su
37	131.5	3.3	1026	1 VG37_BPT4	P03744 bacterioph
38	131.5	3.3	1199	1 N121_RAT	P25291 rattus norv
39	131.5	3.3	2090	1 N214_HUMAN	P35658 homo sapien
40	131	3.2	532	1 SPG7_DICDI	P22698 dictyostell
41	131	3.2	1090	1 GUXB_CELFI	P50899 cellulomona
42	130.5	3.2	1341	1 VG37_BPT2	P07067 bacterioph
43	130.5	3.2	1365	1 GTFS_STRDO	P29336 streptococ
44	130	3.2	1025	1 SLAP_CAUCR	P35828 caulobacter
45	129	3.2	656	1 THIG_SVNY3	Q55710 synechocyst

AC 074623;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Agglutinin-like protein 3 precursor.
 GN ALS3.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RX MEDLINE=98309840; PubMed=9644209;
 RA Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.:
 RT "Candida albicans ALS3 and insights into the nature of the ALS gene
 family.";
 RL Curr. Genet. 33:451-459(1998).
 CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
 CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U87956; AAC39486.1; -;
 KW Cell adhesion; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 1119
 FT DOMAIN 433 792
 FT REPEAT 433 468 1-1.
 FT REPEAT 469 504 1-2.
 FT REPEAT 505 540 1-3.
 FT REPEAT 541 576 1-4.
 FT REPEAT 577 612 1-5.
 FT REPEAT 613 648 1-6.
 FT REPEAT 649 684 1-7.
 FT REPEAT 685 720 1-8.
 FT REPEAT 721 756 1-9.
 FT REPEAT 757 792 1-10.
 FT DOMAIN 399 404
 FT POLY-THR. 404 455
 FT DOMAIN 450 557
 FT POLY-THR. 557 563
 FT DOMAIN 593 597
 FT POLY-THR. 630 635
 FT DOMAIN 666 671
 FT POLY-THR. 702 707
 FT DOMAIN 738 743
 FT POLY-THR. 774 777
 FT DOMAIN 1044 1047
 FT POLY-THR. 1047 1047
 FT CARBOHYD 471 471
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 543 543
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 579 579
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 651 651
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 687 687
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 723 723
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 759 759
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 845 845
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 987 987
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1050 1050
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1061 1061
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1119 AA; 119927 MW; 6A3FB3FC8C879A71 CRC64;

Query Match 4.2%; Score 167.5; DB 1; Length 1119;
 Best Local Similarity 18.7%; Pred. No. 0.022;
 Matches 170; Conservative 98; Mismatches 340; Indels 303; Gaps 42;
 QY 7 TWSNVAIGGGFVDGVFNAGPILYVTDIGG-----MYRWDAAANGRW--IPL-LDWV 58

Db 264 TYKNVAGTRPFDVAYI-----SAIDVNSYTLSTANEYTCAGGYWQRPFTLRWT 313
 QY 59 GWNN--WGYNGVVSIAADPINTNKWAAVGYMTNSWDPDNDGAILRSSDQGATWQI--TLP 115
 Db 314 GYRNSDAGSGIVIVATRTVTDSTTAVT---TLPDPN-----RDKTKTIEILKPIP 363
 QY 116 -----FKLGMNPGMGGERLAVDPNDNNDNIIYFGAPSKGLWRSDSCAT----- 160
 Db 364 TTTITTSYGVVTSYSTKTAPIGATVIV---DIPHTTTTTSKWTGCTITSTTHNP 420
 QY 161 -----WSQMTNFPDVGTVIANP--TD-----TTG 182
 Db 421 TDSIDTVIVQVPSNPNTVTTEYWSQ--SFATTTTITGPGNTDVLIREPPNHTVTTE 478
 QY 183 YQSDIOGVVWVAFKSSSLGQASKTIFGVADPNPV-----FWSRD--GGATWQAVPG 235
 Db 479 YMSE-----SYTTSTETAPPGTDSVIIEKPPNPVTTEYWSSESYSYTTTFTAPP 531
 QY 236 APTGFIPHKGVDPVNHVL-----YIATSNCTGPPYDG-----SSGDV 272
 Db 532 GTDSVI-----IKREPPNHTVTTEYWSQSYTTTITVAPPGGTDFVLIREPPNHTVTTE 587
 QY 273 WKFSVTSGTWTRISPVSTD-----TANDYFGYSGLT-----ID 306
 Db 588 WSQSYTT-TTTVIAPPGGTDSVIREPPNPVTTEYWSQSYATTTITAPGETDVL 646
 QY 307 RQHPNTIMVATQISWN-----PDTIFRSDGGATWTRIDWNTSPNRS 350
 Db 647 REPPNHTVTTE--YWSQSYATTTTITAPGETDVLIREPPNHTVT-----TEYWSQS 699
 QY 351 LRVLDSIAEPLTFGV---QNPVPVPSKLGWMD-----AMADPPNSDRML----- 396
 Db 700 FATTTTVPAPPGGTDFVIREPPNHTVTTEY--WSQSYATTTTITAPGETDVLIREPP 758
 QY 397 -----YGTGATLYATNDLTK-----WDSGGQIHIAPMVK 426
 Db 759 NHTVTTEYWSQSYATTTIAPPGETDVLIREPPNPVTTEYWSQ----- 806
 QY 427 LEETAVNDLISPPSGAPLI-----SALGDLGGT-----HADVTAVPSTIFTSPVFTTG 475
 Db 807 -SYTTATTVTAPPGGTDFVITYDMSSEISSESRPHYNTHTLWSTWVIETKITETS 865
 QY 476 TSVD---YAEIAPSIIVRAGSDPSSQPNDRHVAFTDGGKNWFGSGEGVTTGGTVA 531
 Db 866 CEGDKCCKSWVSVSTRITVTPNNIETPMVTN-----TVDSTTTSTESQSPSGIFSESGVS 919
 QY 532 ASADGSRFVWAPGDGPP-----VYAVGFGNSWAAASQGVPAQAQIRSDRVNPKTFVALS 586
 Db 920 VETESVTVTTAQTNPSVPTTESEVVF-----TTKNNENGPYESPSTNVKSSMD-E 969
 QY 587 NGTFYRSTDGGVTFQPVAAAGLPSGAVGMFHAVPGKEGDLWLAASGLYHSTNGSSWS 646
 Db 970 NSEFTTSTAATSTDIENETIATGVS-----EASSPISSADETTTV 1013
 QY 647 AITGVSSAV-----NVGFGKAPSGSYPAVFFVGTIGVGTGAYRSDDCGTTWVLINDQ 700
 Db 1014 TTTAESTVIEQPTNNNGGKAPSATSSPTTTTANNDVSI-----TGTSTNSQS 1066
 QY 701 HOYGNWQAIT 711
 Db 1067 SQYNSDTQQT 1077
 RESULT 3
 ID N153_RAT
 AC P49791;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa
 DE nucleoporin).

GN NUP153.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 602-613; 622-645 AND 971-993.
RC STRAIN=BUFFALO; TISSUE=Liver;
RX MEDLINE=93137325; PubMed=8422679;
RA Sukegawa J., Blobel G.;
RT A nuclear pore complex protein that contains zinc finger motifs,
RT binds DNA, and faces the nucleoplasm.;
RL Cell 72:29-38(1993).
CC -|- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
CC COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
CC ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC -|- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
CC TERMINAL RING STRUCTURE OF THE NUCLEOLAR CAGE.
CC -|- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -|- PTM: PHOSPHORYLATED IN INTERPHASE, HYPERPHOSPHORYLATED DURING
CC MITOSIS. MAY PLAY A ROLE IN THE REVERSIBLE DISASSEMBLY OF THE
CC NUCLEAR PORE COMPLEX DURING MITOSIS.
CC -|- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUPI, NSP1, POM 121 AND MAMMALIAN P62.
CC -|- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L06821; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR001876; Znf-RanBP.
DR Pfam: PF00641; Znf-RanBP; 4.
DR SMART: SM00547; Znf-RBZ; 4.
DR PROSITE: PS01358; ZF_RANBP2_1; 4.
DR PROSITE: PS01359; ZF_RANBP2_2; 4.
KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding;
KW Phosphorylation.
FT DOMAIN 4 15 GLY-RICH.
FT DOMAIN 442 446 POLY-GLY.
FT ZN_FING 657 687 RANBP2-TYPE 1.
FT ZN_FING 721 750 RANBP2-TYPE 2.
FT ZN_FING 789 818 RANBP2-TYPE 3.
FT ZN_FING 846 875 RANBP2-TYPE 4.
SQ SEQUENCE 1468 AA; 152823 MW; C3DFD9697C556A7C CRC64;

Query Match 4.1%; Score 166; DB 1; Length 1468;
Best Local Similarity 21.8%; Pred. No. 0.036;
Matches 144; Conservative 71; Mismatches 259; Indels 186; Gaps 31;

QY 106 GATWQITPLP--FKLGNMPPGRNGERLAVDPNNNDILYFGAGSGKLMRSTDSGATWSQ 163
DB 884 GTSSSLNPAPSAFKFGIPSSSSGLSTQTTSTGN---FREGDGGKFLGTSSDSGSTNTM 939
QY 164 MTNF-----PD-----VGTYIANPTDTFTGYQSDIQGVV 191
DB 940 NTNFKPKPTGDKFGVLPLDPSKPEIKNDKNDNFQFPPSSGLSNPASSAPFQF---GVS 996
QY 192 WVAFDKSSSSLSQASKTIF---GVADPNPNVFWSDGGATWQAVGAPTGFPHKGVFD 248
DB 997 TLGQOQKEELPOSSAGSFSGAGVANPSSAAI---DVTTSNKNKSGFNFTDITKSV-- 1051
QY 249 PVNHVLYATNTGPGVDDSGDGVKWFVSTGTWTRISVPVPSDTDANDYFGYSGLTIDR- 307
DB 1052 -----SVTPFYKTEAKKEDASATKGGFT-FGKVDSSAALSPSFVLGRTEKQ 1100
QY 308 QHPNTIMVATQISWPPDTIIF-RSTD-----GGATWTRIDWTSTYPNRSLRYV 354

DB 1101 QEPVT-----STSLVFGKKADNEPKCPQVFSFGNSEQTKD----- 1136
QY 355 LDISAPFWLTFGQVPPNPVSPKLGWDEAMADPNNSDRMLYGTGATLYATNDLTKWDS 414
DB 1137 -ESSKPTFTFSVAK-----PSVK-----ESDLAKATFAFGQTNTTT--DQ 1176
QY 415 GGGIHTAPMVKLEETAVNDLISPPSGAPLISALGDL-GGFTHADVTAVSTIF---TSP 470
DB 1177 GA---AKPAFSLNSS-----SSSSTPATSSASIFGSSSTSSSPVAAFFVFGQASNP 1227
QY 471 VFTT--GTSVDYAEIIMPSTIIVRAGSDPSPQPNDRHVA-----FSTDGKKNMFGSGEPGV 524
DB 1228 VSSSAFGNSAESSTSQPLFPDQK--PATTSASAAPPFVFGTGASSN---STVSSGF 1282
QY 525 TTGCTGAASADGSRFVWAPGDCQPVVYAVG-----FGNSWAASQGVPANAIQRSDRVN 578
DB 1283 TFGATTSSSSGFFVFGTGHSAFASPAFAGNQTTFGQSGASQ-----PN 1330
QY 579 PKTFYALNSGTFTYRSDGGVTQFQVAAAGLPSSGAGVGMFHAVPGKGDLMWLAASSGLYHS 638
DB 1331 PPSFGSISSSTALFSA-GSQPVPPPTFGTVSSSSQPPVFGQOPSQS-----AFSGGTANA 1384
QY 639 TNGSSWSAITGV-----SSAVNVGFGKSAPG-----SSYPAVVYVGTIG 678
DB 1385 SSVFQFGSSTTFNFTNNPNSGVFTFGASPSTPAAAAQPSGGSGFSGFSPASFTVGSNG 1444

RESULT 4
QUA_AICA STANDARD; PRT; 809 AA.
ID QUA_AICA AC Q59086;
AC Q59086;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Quinate/shikimate dehydrogenase [Pyroloquinoline-quinone]
DE (EC 1.1.99.25) (NAD(P)-independent quinate dehydrogenase).
GN QUA.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.; AND CHARACTERIZATION.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=95095936; PubMed=8002591;
RA Elsemore D.A., Ornstom L.N.;
RT "The pca-pob supraoperonic cluster of Acinetobacter calcoaceticus
RT contains quiaA, the structural gene for quinate-shikimate
RT dehydrogenase.";
RL J. Bacteriol. 176:7659-7666(1994).
RN [2]
RP SEQUENCE OF 1-20 FROM N.A.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=96011389; PubMed=75923351;
RA Elsemore D.A., Ornstom L.N.;
RT "Unusual ancestry of dehydratases associated with quinate catabolism
RT in Acinetobacter calcoaceticus.";
RL J. Bacteriol. 177:5971-5978(1995).
CC -|- FUNCTION: CAN ACT EITHER ON QUINATE OR ON SHIKIMATE.
CC -|- CATALYTIC ACTIVITY: Quinate + pyrroloquinoline-quinone = 5-
CC dehydroquininate + reduced pyrroloquinoline-quinone.
CC -|- CATALYTIC ACTIVITY: SHIKIMATE + PYRROLOQUINOLINE-QUINONE = 3-
CC DEHYDROSHIKIMATE + REDUCED PYRROLOQUINOLINE-QUINONE.
CC -|- COFACTOR: PQQ.
CC -|- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; FIRST STEP. THIS PATHWAY
CC ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO
CC PROTOCATHECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE
CC PATHWAY.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- INDUCTION: BY PROTOCATHECHUATE.
CC -|- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.
CC -----

RP SEQUENCE OF 1-936 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-72;
RX MEDLINE-87165765; PubMed-2435707;
RA Kawazu T., Nakanishi Y., Uozumi N., Sasaki T., Yamagata H.,
RA Tsukagoshi N., Uda S.;
RT "Cloning and nucleotide sequence of the gene coding for enzymatically
active fragments of the Bacillus polymyxa beta-amylase.";
RL J. Bacteriol. 169:1564-1570(1987).
RN [2]
RP SEQUENCE OF 689-1196 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-72;
RX MEDLINE-89123046; PubMed-2464578;
RA Uozumi N., Sakurai K., Sasaki T., Takekawa S., Yamagata H.,
RA Tsukagoshi N., Uda S.;
RT "A single gene directs synthesis of a precursor protein with beta-
amylase activities in Bacillus polymyxa.";
RL J. Bacteriol. 171:375-382(1989).
RN [3]
RP SEQUENCE OF 1-776 FROM N.A.
RC STRAIN-ATCC 8523;
RX MEDLINE-87231094; PubMed-2438660;
RA Rhodes C., Strasser J., Friedberg F.;
RT "Sequence of an active fragment of B. polymyxa beta amylase.";
RL Nucleic Acids Res. 15:3934-3934(1987).
RN [4]
RP DISULFIDE BOND, AND MUTAGENESIS OF CYSTEINE RESIDUES:
RX MEDLINE-91215008; PubMed-1827035;
RA Uozumi N., Matsuda T., Tsukagoshi N., Uda S.;
RT "Structural and functional roles of cysteine residues of Bacillus
polymyxa beta-amylase.";
RL Biochemistry 30:4594-4599(1991).
CC -1- FUNCTION: THE PRECURSOR PROTEIN IS PROTEOLITICALLY CLEAVED TO
PRODUCE MULTIFORM BETA-AMYLASES AND A 48 kDa ALPHA-AMYLASE AFTER
SECRETION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
polysaccharides so as to remove successive maltose units from the
non-reducing ends of the chains.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO FAMILY 14 OF
GLYCOSYL HYDROLASES (BETA-AMYLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 13 OF
GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15817; AAA85446.1; -;
DR EMBL; Y00150; CAA68344.1; -;
DR PIR; A29130;
DR PIR; A29108; A29108.
DR PIR; A32251; A32251.
DR HSP; P36924; I892.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR001554; Glyco_hydro_14.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase; 1.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Multifunctional enzyme; Hydrolase; Glycosidase; Signal;
KW Polysaccharide degradation; Repeat.
FT SIGNAL 1 35
FT CHAIN 36 1196 BETA/ALPHA-AMYLASE.
FT DOMAIN 36 454 BETA-AMYLASE.
FT REPEAT 455 558

FT REPEAT 565 668
FT DOMAIN 669 1196
FT DISULFID 118 126
FT ACT_SITE 124 124
FT ACT_SITE 198 198
FT ACT_SITE 118 118
FT MUTAGEN 126 126
FT MUTAGEN 358 358
FT CONFLICT 1 1
FT CONFLICT 67 67
FT CONFLICT 100 100
FT CONFLICT 154 154
FT CONFLICT 177 177
FT CONFLICT 227 228
FT CONFLICT 330 330
FT CONFLICT 425 425
FT CONFLICT 493 493
FT CONFLICT 532 532
FT CONFLICT 559 559
FT CONFLICT 665 665
FT CONFLICT 681 681
FT CONFLICT 725 728
FT CONFLICT 736 736
FT CONFLICT 741 741
FT CONFLICT 758 758
SQ SEQUENCE 1196 AA; 130893 MW; A41EA6B70F257064 CRC64;
Query Match 3.8%; Score 152; DB 1; Length 1196;
Best Local Similarity 21.1%; Pred. No. 0.18;
Matches 167; Conservative 72; Mismatches 282; Indels 270; Gaps 45;
QY 61 NNWG-----YNGVVSIAAD-----PINTNKV-WAAVGMVYTNVDNDGAILRS 102
Db 53 NDWGSFKQLQTLKNNGVYAITTDVWVGYSAGDNQFDMSYKYKYANA-----V 102
QY 103 SDGATWQITPLFKLGNMPPGRGMLAVDPNNDNLYFGAPSGKGLARSTDGATWS 162
Db 103 KEAGLKWPIIITHKCGN-----VGD-----DCN-----IPLPS-----WLSSKGSDEM 143
QY 163 QMTNFPDVGTYIANPTDTTCY-QSDIQGVYVAVFDKSSSLGQASKTIFVGVADPNPNVF 221
Db 144 QF-----KDESGYANSEALSPWLSGTGKQYDELYASFAENFAGYKSIPIKY 190
QY 222 WSRDGGATWQAVGAPTGFT-----PHKGVF-----DPVNHVLYIATSNVTPGP 264
Db 191 LS-----GGPSGELRYPSPYPAAGWSYFGRGKFAQYTETAKNAFRTAMNDKYS 239
QY 265 YDSSGDVWKFVSYTGWTTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQISWPD 324
Db 240 LDKINA-AWGTCLTS--LSQINPPTDGDGYTNGGY-----NSAYGKDFLSWQS 286
QY 325 TI-----IFRSTDG-----GATWTR-----IWDWTSYPNR 349
Db 287 VLEKHLGVICAAAHKNFDSVFGVRIGAKISLGHWMNNPAMPHGTEQAGGYDY----NR 342
QY 350 SLRYVLDISAEPWLT-----GVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGA 401
Db 343 LIQKFKDADILD--LTFTCLEMSDSGTAPNYSLPS-----TLVDTVSS--IANAKGV 389
QY 402 TLVATNDLTKWDSGGQIHAPMVK---GLEETAVNDLISPPSGAPLISALGDLDGFE--- 454
Db 390 RLNGENALPTGGSGFQKIEKITKFGYHGFTLLKRNILVN-NDGSP-----TGELSGFKQY 444
QY 455 -----THADVTAVPSTIFTSPVF-----TTGTSVDYAEIN--PSIIV 489
Db 445 IISKAKPDNNGGTGNKVTIYKGFNSPYIHYRPAAGSWTAAPGVKMQDAEISYAKITV 504
QY 490 RAGSFDPSQPNDRHFAVSTGDKGNWFQGS-----EPGGVTGGTGVAAADG 536
Db 505 DIGS-----ASOLEAAFN-DGNNNDNSNTKNYSFSTGTSTYTPGNSGNAGTITSGAPA 557

QY 537 SRFWAPGDPG-----OPVVYAVGF-----GNSWAASGVGPAN-----AQIRS 574
 Db 558 GA---NPGDGGTNTKVVYVYKGFNSPYTHYRPAAGSSWTAAPGVKMQDAEISGYAKITV 614
 QY 575 DRYNPKTFYAL-----SNCT-FYRSTDGVTQPVAAAGLPSSCAVCVMFHAVPGKE 624
 Db 615 DLSASQLEAFAFNDGNNNDNNKTNKFLSTGVTYTP-----GNGAAGTIRTGAP--S 667
 QY 625 GDLWLAASSGLY-----HSTNGGWSAITGVSSAVNVGFGKSAPGSSYPFAVVGVTGGV 680
 Db 668 GSV-LSVVTSTYATDLNEVTGPIQTEKLSGV--LNVSTSTYAPNSN--GVEYTAQTEAP 722
 QY 681 TGAIRSDDCGT 691
 Db 723 SGAFSTMDLGT 733

RESULT 7
 PGCA_MOUSE STANDARD; PRT; 2132 AA.

AC 061282; Q64021;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN AGC1 OR AGC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Cartilage;
 RX MEDLINE=95104847; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fuelep C., Horvath P., Doege K.J., Glant T.T.;
 RT "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan.";
 RL Genomics 22:364-371(1994).
 RN [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.;
 RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene.";
 RL Nat. Genet. 7:154-157(1994).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION IN THE G1 DOMAIN).
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L07049; AAC37670.1; -
 DR EMBL; S73722; AAB32160.1; -
 DR EMBL; S73721; AAB32160.1; JOINED.
 DR HSSP; P98066; LTSG.
 DR MGD; MGI:99602; Agc.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXXSG.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR InterPro; IPR001304; lectin_C.
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF02339; SGXXSG; 62.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00406; IGV; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS01241; LINK; 4.
 DR PROSITE; PS00615; C-TYPE_LLECTIN_1; 1.
 DR PROSITE; PS50041; C-TYPE_LLECTIN_2; 1.
 KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 GI-A.
 FT DOMAIN 152 247 GI-B.
 FT DOMAIN 253 349 GI-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 587 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.
 FT DOMAIN 1917 2132 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 580 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1922 1933 BY SIMILARITY.
 FT DISULFID 1950 2042 BY SIMILARITY.
 FT DISULFID 2018 2034 BY SIMILARITY.
 FT DISULFID 2049 2092 BY SIMILARITY.
 FT DISULFID 2078 2105 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 1171 1173 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDFC6CDBA163 CRC64;

Query Match 3.7%; Score 149; DB 1; Length 2132;
Best Local Similarity 19.5%; Pred. No. 0.53;
Matches 163; Conservative 96; Mismatches 286; Indels 292; Gaps 39;

QY 64 GYNGVSIADPINTNKVAAVGYNTSWDPNDGAILRSSDQA-----TWQITPL 114
DB 1105 GYVSGIPSGDGTET-----SASVEDVSGLPSSGEGLETSASVEDLGPSTRSLETSAS 1160
QY 115 PFKLGGMPGRGMRGLAVDPNDNINILYFGAPSGKGLWRSTDSGATWSQMTNFPD----- 169
DB 1161 GVDVTFPPSGRGPETSVSGVGD-----FSLGPSKGLETSASGA--EDLSGLPSGKEDL 1215
QY 170 -----VGYTIANPT-DITGQSDIQGVVWVAFDSSSLGQASKTIFVGA 214
DB 1216 VGSASGALDFGLKPLPCTGLSGQTPVNGFPSPGSG-----EYSGADIGSGPSS--GLP 1266
QY 215 DPNNPVFWSRDGATQWQAVPGATGTPHKGVPDPVNHVLYIATSNTPGPGYD----- 267
DB 1267 D-----FSLGPSG--PTVSLVD--STLVEVITATSSLEGRTIGIS 1306
QY 268 SSGDVVKFSVTGTRISPTSPSTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTII 327
DB 1307 GSGEV-----SGL-----PLGELSSAD--ISGL-----PSGTLSQASGSPDS-- 1344
QY 328 FRSTDGATWTRIDWTSYPNRSRLVLDISAEPWLTFTGVPQNPFPVPSPLKGMWDEMAI 387
DB 1345 SGETSG-----FFDVSGQPPFGSGVYS-----EETSGI 1371
QY 388 DPNSDRMLYGTATLYATNDLTKWD--SGQIHIAPMWKGL-----EET 430
DB 1372 PETSGQPS--GTPTT-ATSGVTELNSGQDPVSDGSGIIFGSGQSGITSVSGETS 1428
QY 431 AVNDLISPPSGAPL-----ISALGDLGGFTHADVTAVPSTFT----- 468
DB 1429 GISDLSQSPGFPVFGCTATRTPLASGTISGSGESSGITFVDTSFVEVPTTFREBEGL 1488
QY 469 -----SPVFTGTSDYAEALNPISIIIVRAGSFDPSQPNDRHVAFTDGGKNWF 516
DB 1489 GSVELSGFPGSETGELSGTSDVDVSSQSGAIDSSGLTSTPTE-----F 1532
QY 517 QGSEPG-----GVTTGGTVAASA-DGSRFVWAPGDP-----GQPVVY 552
DB 1533 SGLPSGVAEVSGEFSGVETGSSLPSSGAFDGSGLV--SGFTVSLVDRTLVESITQAPTAQ 1590
QY 553 AVFGNSW-----AASQGV-----ANAIQRSDRVNPKTFYALSNGTFFYR 592
DB 1591 EAGEGPGSILEFSCAHSCTPDISELGSGLDLSLTQSGQMETSTETPSSPY--FSGDFSS 1648
QY 593 STD-GGVTFQVPAAGLPSSGAVGMFHAVPKGCDLWLAASSGLYH----- 637
DB 1649 TTDVSGESIAAATGSGESSGLPEVTLNTSELVEGTEPTVQSLGHGSPMTYISRLSEAS 1708
QY 638 --STNGSSWSAIT-----GVSAV-----NVGFGKAPGSSYPVAFVWGTIGG 679
DB 1709 GDASAGDLGAVTNFPFGSGVSPVPEASSDLSAYPAGVGVSAAPKSLSEFFDLHG 1768
QY 680 VTGAYRSDDCGTT-----WVLINDQHOYGNWGOAITGDHANLRVYIGTNG 726
DB 1769 ITSAFHTDLEMTTPSTEVNSNPTF--QEGTREGSNAPEVSGESSITSDIDGTSG 1823

RESULT 8
ID ISOA_PSEAY
AC P10342; PRT; 776 AA.

DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Isoamylase precursor (EC 3.2.1.68).
GN IAM.
OS Pseudomonas amyloclavata.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=32043;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SB-15;
RC MEDLINE=88243808; PubMed=3379068;
RA Anemura A., Chakraborty R., Fujita M., Nouri T., Futai M.;
RT "Cloning and nucleotide sequence of the isoamylase gene from
Pseudomonas amyloclavata SB-15.";
RN J. Biol. Chem. 263:9271-9275(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JD210;
RX MEDLINE=91064385; PubMed=2248978;
RA Chen J.H., Chen Z.Y., Chow T.Y., Chen J.C., Tan S.T., Hsu W.H.;
RT "Nucleotide sequence and expression of the isoamylase gene from an
isoamylase-hyperproducing mutant, Pseudomonas amyloclavata JD210.";
RN Biochim. Biophys. Acta 1087:309-315(1990).
RN [3]
RP SEQUENCE OF 744-776 FROM N.A.
RC STRAIN=SB-15;
RX MEDLINE=89327147; PubMed=2753857;
RA Anemura A., Fujita M., Futai M.;
RT "Transcription of the isoamylase gene (iam) in Pseudomonas
amyloclavata SB-15.";
RN J. Bacteriol. 171:4320-4325(1989).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=98387895; PubMed=9719642;
RA Katsuya Y., Mezaki Y., Kubota M., Matsuura Y.;
RT "Three-dimensional structure of Pseudomonas isoamylase at 2.2-A
resolution.";
RN J. Mol. Biol. 281:885-897(1998).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
linkages in glycogen, amylopectin and their beta-limits dextrins.
CC -!- INDUCTION: BY MALTOSE.
CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
KNOWN AS THE ALPHA-AMYLASE FAMILY. ISOAMYLASE SUBFAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: J03871; AAA25854.1; -;
DR EMBL: X13378; CAA31754.1; -;
DR PIR: A28109; A28109.
DR PDB: 1BF2; 12-AUG-98.
DR InterPro: IPR000461; Alpha-amylase.
DR InterPro: IPR004193; isoamylase_N.
DR Pfam: PF00128; alpha-amylase; 1.
DR Pfam: PF02922; isoamylase_N; 1.
KW Hydrolyase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 776 ISOAMYLASE.
FT ACT_SITE 401 401 BY SIMILARITY.
FT ACT_SITE 481 481 BY SIMILARITY.
FT ACT_SITE 536 536 BY SIMILARITY.
FT DISULFID 410 422
FT DISULFID 546 616
FT DISULFID 738 766
FT CONFLICT 8 8 A -> G (IN REF. 1).
FT CONFLICT 126 126 F -> C (IN REF. 1).
FT CONFLICT 169 171 GAS -> AH (IN REF. 1).

RESULT	9
ISOA_PESP	


```
Db 302 QAMVQAFHNAIKYKMDV-----VYNHFAEGTWTSSDPTTATYISWGLDN 348
QY 292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWMPDTIIFRSDGGATWRIW 341
Db 349 TTYVELTSGNOYFYDNTGIGANTENTVAQNILVDSLAYWNTM---GVDGFR-----F 400
QY 342 DWTSP-YPNRSRLRYVLDISAEPWLTFGVQPNPPVPSKLGHWMDAAMD-----PFNSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCPNGYNFDAADSNVAINRLREFTVRPAA 451
QY 397 YGATGATYATNDLTKWDGSGQIHF-IAPMVKGLLE-----TAVNDLISPPSGAPLI 445
Db 452 GGSGLDLFA---EPWAIGNSYQLGGFPQGWSEWNGLFRDSLRAQNEL-----GSMTI 502
QY 446 SALGDLGGFTHADVAVPSTFTSPTTGTSTVDYAEIPLSIIVR-----AGS 493
Db 503 YVTQDANDESGS-----SNLFOSGSRSPWNSINFDVHDGMDLKVYSCGANNNSQAWP 556
QY 494 FDPSSQPNDRHVAFTDGGK---NMFQSGEPGVTTGGTV---AASADGSRFVWAPGDP 546
Db 557 YGFS-----DGGSTNYSWDQGMASG---TGAADVQRRRAARTGMAFEML---SA 599
QY 547 GQPVVAVGFGNSWAASQGVPAQAQIRDRVNPKEF-YALSNGTFFYRSTDCGVTFQPVAA 605
Db 600 GTPLMQG---GDEYLRFLQCNNAYNALDSSANWLTYSWTTDQSNFYFAORLIATFRKAHP 656
QY 606 GL-PSSGAVGVMFHAVPKEGDLWLAASSGLYHSTNG-----GSSWSAIT---GVSSAVNVG 658
Db 657 ALRPSS-----WYSGSLTWYQPSGAVADSNVWNTSNYATAYAIN-- 697
QY 659 FGKSAPGSSYPVAVVGTIGVGTGAYRSDCGTTWVLINDDQHQYGNWGAITGDHANLR 718
Db 698 -GPSLGSNSIYYAYNGWSSSFTTLPAPPSTQWRYVTD-----CDWMDGASTFEVAPGS 752
QY 719 RVVIGTNGRIGIVYGDIG 735
Db 753 ETLIG--GACTTYGQCQ 767

RESULT 10
BIGA_SALTY
ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Putative surface-exposed virulence protein biga precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=14028;
RC "Salmonella typhimurium rhs homolog.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RL Nature 413:852-856(2001).
RN [3]
RN SEQUENCE OF 1-765 FROM N.A.
```

```
RC STRAIN=LT2;
RX MEDLINE=911100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting shiroheme
RT cofactor.";
RL J. Bacteriol. 173:325-333(1991).
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF133696; AAD39458.1; -
DR EMBL; AE008859; AAL22340.1; -
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; bigA.
KW Virulence; Repeat; Signal; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1953 PUTATIVE SURFACE-EXPOSED VIRULENCE
FT PROTEIN BIGA.
FT DOMAIN 101 252 15 X 11 AA TANDEM REPEATS.
FT REPEAT 101 103 1 (INCOMPLETE).
FT REPEAT 104 113 2 (INCOMPLETE).
FT REPEAT 114 122 3 (INCOMPLETE).
FT REPEAT 123 133 4.
FT REPEAT 134 144 5.
FT REPEAT 145 155 6.
FT REPEAT 156 166 7.
FT REPEAT 167 177 8.
FT REPEAT 178 188 9.
FT REPEAT 189 199 10.
FT REPEAT 200 210 11.
FT REPEAT 211 221 12.
FT REPEAT 222 232 13.
FT REPEAT 233 243 14.
FT REPEAT 244 252 15 (INCOMPLETE).
FT CONFLICT 207 207 D -> DRGDDVTPDD (IN REF. 1).
FT CONFLICT 514 514 A -> R (IN REF. 3).
FT CONFLICT 1698 1698 D -> N (IN REF. 1).
FT CONFLICT 1795 1798 OYLE -> ITLO (IN REF. 1).
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 3.7%; Score 148.5; DB 1; Length 1953;
Best Local Similarity 20.5%; Pred. No 0.51;
Matches 158; Conservative 83; Mismatches 274; Indels 257; Gaps 34;

QY 44 WDAANG-----RWIPLLDVWGNMNGYGVVSTADPINTNKVWAAVGYMTNSWD---P 94
Db 1344 WNTATGVINGINPDAPLINLGRYNGFNQGTINVGQD---NAV--AISGGTSSYVINLV 1397
QY 95 NDGAILRSSDQAGTWITPLPFLKGNMGP-----RGMGERLAVDPNNDNILYFCAPSGKG 150
Db 1398 NSGTINVTGEQK-----EDGTNGTGLIGKGNATNTTNTADGVINYYADDSYA 1448
QY 151 LWRSTDSGATWSMTNPPDVTGVTIANPTDTGTQSDIQGVVWVAFDKSSSLGQASKTIF 210
Db 1449 FGKTRAIINNGEINLLCDSGCDIYAP-GTGTQNDHNGTADIVIPDATTAPTEGS--IP 1505
QY 211 VGVADPNPFWSRDGGATQWQVPGAPTGFPHKGVFDPVNVHLYIATSTNGPYDSSG 270
Db 1506 TTPADPNAPQOLSN-----YIVGTNA---DGSSG 1531
```


QY 271 DVWKFV-----TSGTWTRISVPSTDTANDYFYCSGLTIDRQHPNTIMVATQ 318
 Db 1532 TLKANLVIGNVKVDGFTSGTADTTVVVDNAFTGSNIQADNIT-----STS 1580
 QY 319 ISWPPDIIIFRSTDDGATWTRINDWTSYPNRSLRYVLDISAEPWLTFTGVQPNPVPSPKL 378
 Db 1581 VVW--NAQGSQDADGNV-----DVTMTKNAYADVATDSSVS----- 1614
 QY 379 GWDMAAIDPFNSDRMLY-----GTGATLYATNDLTKWDSGGQ----- 417
 Db 1615 ---DVAQALDAGYTNNELYTSNLVGTITAEI---NSALKQVSGQAQTTVFREARVLSNRF 1668
 QY 418 --THIAPMVK--GLEETAENVLDLISPPSGAPLISALDGLGFTHADVTAVPSTIETSPVFT 474
 Db 1669 MLADAAFOIKDGLAFNV-----AKGDP---RAELGNDTQYDMLALROTLDLTA--SQ 1716
 QY 475 GTSVDYAEALNPSIIVRAGSDPSPQPNDRHVAFSTDGKKNWFQSGPGVTTGGTVAASA 534
 Db 1717 NLITLEY-----GIARLDG 1729
 QY 535 DGSRFVWAPDGPVYVAVGFG--NSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYR 592
 Db 1730 DGSK---TAGDNGLTGYSOFFGLKHSMAFDEGLAWNLSRYDVHNLDSRSVAYGDVVK 1786
 QY 593 STD-----GGVTFQPVAAGLPSSGAVGVNF--HAVPG--KE---GDLWLAASS 633
 Db 1787 IADSDMRQOYLEFRSEGAFTTMMGDALKVTPYAGVKFRHTMEDGYKERSAGDFNLMSNS 1846
 QY 634 -----GLXHSITGSSWSAITGVSSAVNVGFKS-----APGSSYPAPFVVG 675
 Db 1847 GNETAVDSIVGLKLDYAGKGDWSATATLEGPNLSYSKSORTASLOGAAGQSF--GVDDGQ 1905
 QY 676 TIGVTC-----AYRSDCGTTLWLNIDQHQVGNMQAITGDHANLRVY 721
 Db 1906 KGGVGNLATIGVYKSSND-----TALHLDAYQWKEDGISDKGFMLNVKTF 1952
 RESULT 11
 PGCA_CHICK STANDARD; PRT: 2109 AA.
 ID AC P07898: Q90991; Q90820; Q91047; Q90810;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN Gallus
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LEHORN; TISSUE-Embryo;
 RX MEDLINE=94043149; PubMed=8226878;
 RA Li H., Schwartz N.B., Vertel B.M.;
 RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia.";
 RT J. Biol. Chem. 268:23504-23511(1993).
 RL [2]
 RN SEQUENCE OF 1042-1559 FROM N.A.
 RP TISSUE-Embryo;
 RC MEDLINE=90307744; PubMed=1694853;
 RX Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
 RA "Chick cartilage chondroitin sulfate proteoglycan core protein. II. Nucleotide sequence of cDNA clone and localization of the S103L epitope.";
 RT J. Biol. Chem. 265:12088-12097(1990).
 RL [3]
 RN SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
 RP TISSUE=Cartilage;

RX MEDLINE=93111968; PubMed=1339285;
 RA Chandrasekaran L., Tanzer M.L.;
 RT "Molecular cloning of chicken aggrecan. Structural analyses.";
 RL Biochem. J. 288:903-910(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=94107258; PubMed=8280087;
 RA Chandrasekaran L., Tanzer M.L.;
 RL Biochem. J. 296:885-887(1993).
 RN [5]
 RP SEQUENCE OF 1492-1610 FROM N.A.
 RC STRAIN-WHITE LEHORN; TISSUE=Chondrocytes;
 RX MEDLINE=95128519; PubMed=7827752;
 RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
 RT "Molecular basis of nanomelia, a heritable chondrodystrophy of chicken.";
 RT Matrix Biol. 14:297-305(1994).
 RN [6]
 RP SEQUENCE OF 1894-2109 FROM N.A.
 RX MEDLINE=89008500; PubMed=3170613;
 RA Tanaka T., Har-El R., Tanzer M.L.;
 RT "Partial structure of the gene for chicken cartilage proteoglycan core protein.";
 RL J. Biol. Chem. 263:15831-15835(1988).
 RN [7]
 RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
 RX MEDLINE=86259736; PubMed=3460082;
 RA Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;
 RT "Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT (CHONDRODYSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS. AGGECAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGECAN/VERSCAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L21913; AAB19128.1; -
 DR EMBL; M38187; AAA48731.1; -
 DR EMBL; M88101; -; NOT_ANNOTATED_CDS.
 DR EMBL; S74657; AAC60751.1; -
 DR EMBL; S74656; AAC60751.1; JOINED.
 DR EMBL; J04028; AAA48719.1; -

DR EMBL; M13993; AAA48720.1; --
 DR PIR; A25442; A25442.
 DR HSP; P08709; IBF9.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXSG.
 DR InterPro; IPR000436; Leshi_SCR_CCP.
 DR InterPro; IPR001304; lectin_c.
 DR Pfam; PF00008; EGF_1.
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF02339; SGXSG; 56.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECR; 1.
 DR SMART; SM00179; EGF_Ca; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; Link; 4.
 DR PROSITE; PS0010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01187; EGF_Ca; 1.
 DR PROSITE; PS01241; Link; 4.
 DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE; PS00615; C-TYPE_LECTIN_2; 1.
 DR PROSITE; PS0041; C-TYPE_LECTIN; 1.
 KW Glycoprotein; Cartilage; proteoglycan; Lectin; Signal; Sushi;
 EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 2109 AGGRECAN CORE PROTEIN.
 FT DOMAIN 44 136 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 166 243 LINK 1.
 FT DOMAIN 264 346 LINK 2.
 FT DOMAIN 537 614 LINK 3.
 FT DOMAIN 635 716 LINK 4.
 FT DOMAIN 1363 1742 LINK 4.
 FT DOMAIN 1855 1892 EGF-LIKE.
 FT DOMAIN 1901 2019 C-TYPE LECTIN.
 FT DOMAIN 2023 2081 SUSHI.
 FT DOMAIN 48 137 GI-A.
 FT DOMAIN 148 243 GI-B.
 FT DOMAIN 249 346 GI-B'.
 FT DOMAIN 519 613 G2-B.
 FT DOMAIN 620 715 G2-B'.
 FT DOMAIN 718 803 KS.
 FT DOMAIN 805 1264 CS-1.
 FT DOMAIN 1265 1742 CS-2.
 FT DOMAIN 1893 2109 G3.
 FT DISULFID 51 129 BY SIMILARITY.
 FT DISULFID 171 242 BY SIMILARITY.
 FT DISULFID 195 216 BY SIMILARITY.
 FT DISULFID 269 345 BY SIMILARITY.
 FT DISULFID 293 314 BY SIMILARITY.
 FT DISULFID 542 613 BY SIMILARITY.
 FT DISULFID 566 587 BY SIMILARITY.
 FT DISULFID 640 715 BY SIMILARITY.
 FT DISULFID 664 685 BY SIMILARITY.
 FT DISULFID 1859 1870 BY SIMILARITY.
 FT DISULFID 1864 1879 BY SIMILARITY.
 FT DISULFID 1881 1890 BY SIMILARITY.
 FT DISULFID 1897 1908 BY SIMILARITY.
 FT DISULFID 1925 2017 BY SIMILARITY.
 FT DISULFID 1993 2009 BY SIMILARITY.
 FT DISULFID 2024 2067 BY SIMILARITY.
 FT DISULFID 2053 2080 BY SIMILARITY.
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1856 1856 MISSING (IN ISOFORM 2).
 FT CONFLICT 362 362 E -> D (IN REF. 3).
 FT CONFLICT 601 601 G -> D (IN REF. 3).
 FT CONFLICT 1000 1000 P -> R (IN REF. 3).
 FT CONFLICT 1029 1029 A -> P (IN REF. 3).
 FT CONFLICT 1043 1043 VT -> PA (IN REF. 2).
 FT CONFLICT 1251 1251 E -> D (IN REF. 2 AND 3).
 FT CONFLICT 1587 1587 I -> T (IN REF. 5).
 FT CONFLICT 1590 1590 I -> V (IN REF. 5).
 FT CONFLICT 1594 1594 T -> S (IN REF. 5).
 FT CONFLICT 1602 1610 IETSTVREI -> VLRCRSVLR (IN REF. 5).
 FT CONFLICT 1603 1603 E -> A (IN REF. 3).
 FT CONFLICT 1672 1672 S -> G (IN REF. 3).
 FT CONFLICT 1796 1796 E -> G (IN REF. 3 AND 7).
 FT CONFLICT 1988 1988 F -> S (IN REF. 6).
 SQ SEQUENCE 2109 AA; 223492 MW; 7F824FDB53A2ABDA CRC64;
 Query Match 3.6%; Score 146; DB 1; Length 2109;
 Best Local Similarity 20.5%; Pred. No. 0.78;
 Matches 147; Conservative 72; Mismatches 250; Indels 248; Gaps 32;
 QY 94 PNDGAILRSSDQATWQITPLPKLGNMPCRMGERLAVDPNNDNL--YFGAPSGKGL 151
 DB 818 PTSGEPSASGAPDTSGETSVGFELSGEQSGIGESGLPSVDLQSLGSGFVPGESGLPSG--- 874
 QY 152 WRSTDGATWSQMTNPDVCTYIANPTDTTGYQSDTQGVVWVAFDKSSSSLSGOASKTIFV 211
 DB 875 -----DVSLPSGIVDI-----SCLPAAEEVTV 898
 QY 212 GVADPNPNVFWSDGGATWQAVPGAPTGFPHKGVDPVNVHLYIATNTGGPYDGGSGD 271
 DB 899 SV-----SRIPEVSGMPSG-----AESGLHSGFSGE 925
 QY 272 VKFSTVSGTWTRISVPSTDTANDYFGYSLRIDRQHPNTIMVAT----- 317
 DB 926 I-----SGT-ELISGLPSGEESGLASGFP--TISLVDSTLVEVVTAAAPGRQEGKSG 976
 QY 318 -----QISWMPDIIFRSTDGGATWTRIMDWTSPYKNSRLRYLVDISAEPWLTGVPQNP 372
 DB 977 VSCEEELSGFPSA-----EWDSSGARG-----LPSGAE--TSGEQSGVP 1013
 QY 373 -----VSPKXGLWDEAMAIIDPNFNSDRMLYGTGATLYATNDLTKWDSG-----GQIH 419
 DB 1014 ELSGEHSQVPSA---GLSGEAFVPELSGBH---SGVT-----ELSGEHSGLPELSGEPF 1060
 QY 420 IAPMWGLEETAVNDLISPPSGAPLISALDGLGFT-----HADVTAVPSTITSPVFTT 474
 DB 1061 GVPELSGFPSSL--DISGEPGAPVSGVPDVSLTSGVDGSGEVSGV-TFISTLQEV 1117
 QY 475 GTSVDYVAELNPSIIVR-----AGSFDPSOPNDRHVAFTDGGKWNWFGSGEPGG 523
 DB 1118 TPSVAEAEAKEILEISGLPSGETSGMVSGSLDVSGQPSG-HIGF-----GGSASG 1166
 QY 524 VT-----TGCTVAASADGSRFVWAPDGPQPVVYAVYVFGNSAASQGVAPANAQIRSDRV - 577
 DB 1167 VLEMSGFPFGAVESGSEAS-----GVEVTSGLASGEESGLTSGFPTVSLVDITLVE 1217
 QY 578 -----NPKTFYALSNGTFYRSTYDGGVTPQVAAAGLPSSGAVGVNMFHVPCKE 624
 DB 1218 VVTQTSVAQEVGEGPSCMIIEISG---FLSGDRGVS-----GEGSGAVQSSGLPSGT 1265
 QY 625 GDLWLAASCLYHSTNGCSSWSAITGVSSAVNVVFGCKAPGSSYPVAVVGT--IGGV 682
 DB 1266 GD-FSGEPGIPYFSGDISGATDLSGQPSAVT---DISGEDSGLPVTLVSDLVVVTR 1321
 QY 683 AYRSDDCGTTWVLINDQHQYGNWQAITGDHANLRVYIGTNGRIGVYDIDGAPS 739
 DB 1322 PTYSQELG-----GETAVTFYVFGPSGEGSASGDLGGAS 1357

Db 585 EV--YRGSTLIVARTGSHVTGLSAAATYFTTAVRAVDAAGNVS-AASAPVGVTTAPDPT 641

QY 665 -GSSYPVAVVGTIGGVTGAYRSDCGTT----WVL 695

Db 642 TGSCAVTYTANGSGGFTAAVTLTNVTTALSGWTL 677

RESULT 15

NUP98_RAT

ID NUP98_RAT STANDARD; PRT; 937 AA.

AC P49793;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Nuclear pore complex protein Nup98 (Nucleoporin Nup98) (98 kDa

DE nucleoporin).

GN NUP98.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Liver;

RX MEDLINE=95254630; PubMed=7736573;

RA Radu A., Moore M.S., Blobel G.;

RT "The peptide repeat domain of nucleoporin Nup98 functions as a

RT docking site in transport across the nuclear pore complex.;"

RL Cell 81:215-222(1995).

CC -I- FUNCTION: PLAYS A ROLE IN THE BIDIRECTIONAL TRANSPORT ACROSS THE

CC NUCLEOPORIN COMPLEX (NPC). THE REPEAT DOMAIN HAS A DIRECT ROLE IN

CC THE TRANSPORT.

CC -I- SUBCELLULAR LOCATION: Nuclear pore complex.

CC -I- DOMAIN: CONTAINS G-L-F-G REPEATS.

CC -I- PTM: THE N-TERMINUS IS BLOCKED.

CC -I- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L39991; AAC42054.1; -

DR InterPro: IPR004325; Nucleoporin_FG.

DR Pfam: PF03093; Nucleoporin_FG; 26.

KW Nuclear protein; Transport; Repeat.

FT DOMAIN 889 893 POLY-GLU.

SQ SEQUENCE 937 AA; 97809 MW; 6B1F108C7390307A CRC64;

Query Match 3.6%; Score 143.5; DB 1; Length 937;

Best Local Similarity 20.4%; Pred. No. 0.41;

Matches 137; Conservative 78; Mismatches 222; Indels 233; Gaps 34;

QY 143 FGAP-SGKGLWRSTDGATWSQMTNPDVGTI--ANPTDTTCYQSDIQGVVWVAFDK-- 197

Db 6 FGTPFGTGGFGT---STFGQNTGF---GTTSGGAFGTSAFGSSNTGGLFGNSQTKPG 59

QY 198 ---SSSLGQ-----ASKTIFVGVAADPNPNFVWSRDGGATWQAVPGAPT 238

Db 60 GLFGTSSLSQPAVSTGFGFGTGTGTSNSLF-GTANTGTSLSFSSQNNNA---FAQNKPT 114

QY 239 GFTHPKGVF--DPNVHVLVYATNTGNGPYDGGSGDVWKF-----VTSGTWTTRISPVPTSD 292

Db 115 GF-----GNFCTSSGGLFGTNTTNSPFGNTSCSLFGPSSFTAAPTGTIKENPPTGTD 170

QY 293 TANDYFCYGLTIDRQHPNTIMVATOISWNPDIIFRSTDGGATWTRINDWTSYPNRSRLR 352

Db 171 T-----MVKAGV-----STNISTKHOCITAMKEYESKSLD 200

Search completed: July 2, 2002, 09:18:03

Job time: 371 sec

QY 353 YVLDISAEPW-----LTFGVQPNPPVPSPKLGMMDEAMADPFN--SDRML 396

Db 201 ---ELRLEDYQANKRGPNQVAGAGTTTGLFGSSPATSSATGLFSSSTTNSAFSYGQNKTA 257

QY 397 YGTGATLYATNDLTWKWDSGGQIHIAPMVKGLEETAYNDLISPSPGAPLISALDGLGGFTH 456

Db 258 FGTSSTTGFGTN-----PGG-----LFGQQNQQTTSLSFKPFG----- 289

QY 457 ADVTAVPSTIFTSPVFTTGTSDVYAEINLPSIIVRAGSEDPSSOPNDRHVAFTDGGKNWF 516

Db 290 -QATTPNTGFS-----FNTSTLGPST-----NTWGLEGV 321

QY 517 QGSEPGGVTTGTTVAASADGSRFVWAPGDPQVPVYAVGFGNSWAASQGVYAPANAQRSDR 576

Db 322 QASQPGGLF--GTATNTSTGTAFGTGTLFGQP---NTGFG---AVGSTLFGNKKLT--- 370

QY 577 VNPPTYALSNGTFYRSTDGCVTFQPVAAGLPSSGAVGVMFHAVPKGEGDLWLAASGL- 635

Db 371 -----TFGTSTTSAPSGFTTSGGL-----FGNKPPTLTLTGNTNTS 405

QY 636 ---YHSTNGGS---SWSAITGVSSAVNVYFGKSAKPGSSYPVAF-----VVGTFIGVGTG 682

Db 406 NFGFGTNNSSSIFGSKPAAGTLGTGLTGFG--TALGAGQASLFGNNQPKIGPLG--TG 482

QY 683 AYRSDDCGTTWVLI-----NDDQHQYGNWQQAITGDHANLRRVYIGTNGRG 728

Db 463 AFGAPGFNTSTAILGFGAPQAPVALTDPNAAQA-----QAVLQQHLN-----S 506

QY 729 IVYGDIGGAP 738

Db 507 LTYSPEGDS 516

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:22 ; Search time 91.42 Seconds
(without alignments)
1400.309 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTPQYTSNVAIGGGFVD.....YIGTNGRGIIVYDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2371	58.7	890	2	O86727	O86727 streptomyc
2	2052.5	51.1	996	2	O9AQH0	O9aqh0 caldicellul
3	2009	49.8	839	16	O97KK0	O97kk0 clostridium
4	1753.5	43.4	806	3	O9P4T8	O9p4t8 agaricus bi
5	1680	41.6	856	3	O74170	O74170 aspergillus
6	1160.5	28.8	707	16	O9WYE1	O9wye1 thermotoga
7	238	5.9	2468	16	O9I2M3	O9i2m3 pseudomonas
8	214	5.3	1800	2	O9L948	O9l948 pseudomonas
9	202.5	5.0	2174	16	O92U08	O92uu8 rhizobium m
10	189	4.7	3972	16	P73139	P73139 synechocyst
11	186.5	4.6	4199	16	P74440	P74440 synechocyst
12	185.5	4.6	818	6	O9N1P0	O9n1p0 bos taurus
13	185.5	4.6	997	2	O9Z4I1	O9z4i1 bacillus sp
14	184.5	4.6	998	17	O9HLQ9	O9hlq9 thermoplasm
15	182.5	4.5	848	2	O9RK65	O9rk65 streptomyc
16	182	4.5	1904	9	Q38319	Q38319 lactococcus

ALIGNMENTS

RESULT 1

O86727 ID O86727 PRELIMINARY; PRT; 890 AA.
AC O86727;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED CELLULOSE.
GN SC5C7_30C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RM Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031515; CRA20642.1; -;
DR HSSP; P07986; IEXG.
DR InterPro; IPR002860; BNR.
DR InterPro; IPR001919; CBD_2.
DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBD_2; 1.
SQ SEQUENCE 890 AA; -93252 MW; D2C58695A4B56E84 CRC64;

17 175.5 4.3 2523 16 O53393
18 174.5 4.3 3930 16 O98E20
19 167 4.1 1804 1 O9P9A9
20 167 4.1 3659 16 Q98LN6
21 166.5 4.1 902 16 Q9AAJ5
22 166.5 4.1 3145 16 Q98MG7
23 166 4.1 699 16 O07219
24 186 4.1 1203 5 Q9N5K0
25 164.5 4.1 695 2 Q934G0
26 164 4.1 631 6 Q9N1P1
27 163.5 4.1 1047 3 O93865
28 163.5 4.1 13055 5 Q09165
29 163 4.0 1684 2 Q03658
30 160.5 4.0 3716 16 O50379
31 159.5 4.0 1915 2 Q9RPL0
32 158.5 3.9 762 2 Q9RM63
33 158.5 3.9 1441 16 Q9CFL1
34 158 3.9 2232 5 P91365
35 156.5 3.9 695 2 Q9F9U2
36 156.5 3.9 2204 16 O07231
37 156 3.9 2117 3 Q96U11
38 156 3.9 3157 16 O50378
39 155 3.8 809 5 Q9NAN8
40 155 3.8 4180 16 Q9I5N6
41 154.5 3.8 3570 4 Q99552
42 154 3.8 779 2 Q938A2
43 153 3.8 1055 16 Q9A9P9
44 153 3.8 1115 4 Q9HD43
45 153 3.8 13288 6 O18758

O53393 mycobacteri
Q98E20 rhizobium l
Q9P9A9 uncultured
Q98LN6 rhizobium l
Q9AAJ5 caulobacter
Q98MG7 rhizobium l
O07219 mycobacteri
Q9N5K0 caenorhabdi
Q934G0 pseudomonas
Q9N1P1 bos taurus
Q93865 candida alb
Q09165 caenorhabdi
Q03658 unidentified
O50379 mycobacteri
Q9RPL0 acetivibrio
Q9RM63 myroides od
Q9CFL1 lactococcus
P91365 caenorhabdi
Q9F9U2 pseudomonas
O07231 mycobacteri
Q96U11 neurospora
O50378 mycobacteri
Q9NAN8 caenorhabdi
Q9I5N6 pseudomonas
Q99552 homo sapien
Q938A2 stigmatella
Q9A9P9 caulobacter
Q9HD43 homo sapien
O18758 sus scrofa


```
Db 142 GNMPCRGMRKLVVDPNLSLYFGARSGNGLMKWSTNGRSWSKVFTFTDGTSGFVDPDSD 201
Qy 180 TTGYSQSDIQGVVWVAFKSSSLQOASKTIFVGVADP--NNPVFWSRDGATWQAVPGAPT 238
Db 202 STGLNSDKIGIAWTIDKASGSSSATPRIFVGVANKGNSIYISTNGGSSWSAVAQOPT 261
Qy 239 GFIPHKGVDFDPVNHVLIATSNTPGPGYDGGSDGVWKFVSFTSGTWTRISPPVSTDTANDYF 298
Db 262 SFLPHKGVLSERALYITVSDGGPGYDGTSGAVYKYSIDTGAWTIDTPVSGSDL--FF 318
Qy 299 GYSLGTTDRQHPNTIMVATQISWPPDIIIFRSTDDGATWTRINDWTSYPNRSRLRYVLDIS 358
Db 319 GFGGLADLTQKNGVWVAALNSWMPDQOIFRSTNGASWTPLDWAVYPTLNKYYISNAA 378
Qy 359 AEPWLTEGVOPNPVPSPP--KLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSG 415
Db 379 LAPW----IGPNIVDVTGPNLQIGMMESLIDPFDSNHWLYGTGETIYIGSROLLKWD 434
Qy 416 GQIHAPWVKGLETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVETTG 475
Db 435 HNVTIKSLADGVEETSQALISPPSGPLVSAIGDVGFGVHQSILTRAPSAQFTNPKWSTS 494
Qy 476 TSVDYAEINPSLIIVRAGSEFPSSQPNDRHVAFTSDGCKNMFQSGSEPGVTTGGTVAA 535
Db 495 ADIDFAGNNSNVVRIGTGDSST---CKQVAISSDYGVWVNHQHFAGDPNVQGGKVAL 551
Qy 536 GSRFVWAPGDPGQPVVYAVFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFYRST 595
Db 552 ADIILWRTNGNG-----VMYSRQATNFIYGASGKTFYVYSTD 589
Qy 596 GGVTFQPVVAGLPSGAGVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSATGVS 655
Db 590 NGKTFSAHSGLSGATSVDFTVH--PVSGBDIWASTDKGLFHSDDSGATFSAISGITQ 647
Qy 656 NVGFGKSAPOSSYPVAVFVGTIGVTCAYRSDCGTWTWVINDDQHOYG--NWGAITGDH 714
Db 648 GVALGAPRSTGGYPVAFVAAANYGEVAYFRSDDRGVWVWVINDAAGHGAASANCMAAD 707
Qy 715 ANLRVYIGTNGRGIVYDGGGA 737
Db 708 RVYGRVYIGTNGRGIFGYDVAGS 730

RESULT 5
ID 0741170 PRELIMINARY; PRT; 856 AA.
AC 0741170;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE AVICELASE III.
GN AVIII.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RT "Avicelase III from Aspergillus aculeatus.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015511; BAA29031.1;
DR HSSP; P00725; 2CBH.
DR InterPro; IPR002860; BNR.
DR InterPro; IPR000254; CBD_fungal.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBD_1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;
```

```
Query Match 41.6%; Score 1680; DB 3; Length 856;
Best Local Similarity 46.2%; Pred. No. 3.4e-86;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

Qy 1 ATTQPTWTSNVAIT--GGGFGVDGIVFNEGAPGILYVRDIDIGMYRWDAANGRWPLLDWVG 59
Db 21 AASQAYTWKNVVTGGGGGFTPGIVFNPISAKGAYARTDIGGAYRLN--SDTWTPLMDWVG 79
Qy 60 ---WNNNGVGVVSIADPINTNKVAAVGYMYSWPNBNGAILRSSDQAGTQITPLPF 116
Db 80 NDTWHDW---GIDALATDPDVTDRVYVAVGMYTNEWDPNVGSIILRSDQGTWTETKLFP 136
Qy 117 KLGNNMPCRGMRGERLAVDPNNNDILYFCAPSGKGLMRSTDSGATWSQMTNPPDPVGTYIAN 176
Db 137 KVGNNMPCRGMRGERLAVDPNKNLSILYFGARSGHGLMKSTDYGATWSNVTSFTWTGTVFQD 196
Qy 177 PTDTTGTQSDIQGVVWVAFDKSSSLQOASKTIFVGVADPNPNPVFWSRDGATWQAVPGA 236
Db 197 SSST--YTSDPVGLIAWVTFDSTSGSSGSATPRIFVGVADAGKSVFKSEDAATWAWVSGE 254
Qy 237 PT--GFIPHKGVDFDPVNHVLIATSNTPGPGYDGGSDGVWKFVSFTSGTWTRISPPVSTDTAN 295
Db 255 PQXGFLPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNIITSGVWTDISP---TSLAS 311
Qy 296 DYFGYSGLTITDRQHPNTIMVATQISWPPDRIIFRSTDDGATWTRINDWTSYPNRSRLRYL 355
Db 312 TYXGYGGLSYDLQVPGTLMVAALNCWMPDELIIFRSTDSGATWSPIWEWNGYPSINYYSY 371
Qy 356 DISAEPWLITGVQPNP--PVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDS 414
Db 372 DISNAPWIQDVTSTDDQPPV--RVGMVVEALAIIDPFDSNHWLYGTGLTVYGGHDLTNWDS 428
Qy 415 GGOIHAPWVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTT 474
Db 429 KHNVTYKSLAVGIEEMAVGLITPPGPGALLSAVDDGGFVHSDLDAAPNQAYHTPYGT 488
Qy 475 GTSVDYAEINPSLIIVRAGSEFPSSQPNDRHVAFTSDGCKNMFQSGSEPGVTTGGTVAA 534
Db 489 TNGIDYAGNKPSNIVRSASD--DYPT---LALSNEFGSTWYADYAAASTSTGTGAVALS 543
Qy 535 DGSREFWAPGDPGQPVVYAVFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFYRST 594
Db 544 DGDVTLLMSSTSGALVSKSG--TLTAVSSLPSGAVIASDKSDNTVYFGSGAGAIIVSK 600
Qy 595 DGGVTFQPVVAGLPSGAGVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAT--TGVS 653
Db 601 NTATSTFKTVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTA 658
Qy 654 AVNVGFGKSAPOSSYPVAVFVGTIGVGTGAYRSDDCGTTWVINDDQHOYGNWGA--IT 711
Db 659 GWSFGFGKASSTGTVYVYIGFTIDGAAGLFKSEDACTNMQVSDASHGFGS--GSANVN 717
Qy 712 GDHANLRVYIGTNGRGIVYDGGAPSG 740
Db 718 GDLQTYGRVFRGHERPHGLLRQSQREPAG 746

RESULT 6
Q9WYE1 PRELIMINARY; PRT; 707 AA.
AC Q9WYE1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOGLUCANASE, PUTATIVE.
GN TW0305.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
```



```
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.B., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001712; AAD35393.1; -.
DR TIGR; TM0305; -.
DR InterPro; IPR002860; BNR.
DR Pfam; PF02012; BNR; 9.
SQ Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CDBA6A99F CRC64;

Query Match 28.8%; Score 1160.5; DB 16; Length 707;
Best Local Similarity 35.2%; Pred. No. 3.2e-57;
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

QY 6 YTSNVAIGGGGFGDGVINEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWGMNNGY 65
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 FEWKSVIEINGGFGVPGIIFHPASGLLYARTDVGGLYRWDEETKRWKQLDFLRRQSDY 80
QY 66 NGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLP---FKLGGNM 122
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 MGVLVALDPSPKRIYATGKYTDW-AGYGAILSEIDYGETWTIVNLDKYGKVGNGE 139
QY 123 PGRGMEERLAVPNDNINILFYGAPSGKGLWRSTDGATWSQNTNFPDVGTYIANPTDTTG 182
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
140 DGRNAGERLQVDPNFSVLFMGIT-KYGLWKSEDFGKNNKKVDSFST----- 186
QY 183 YQSDIOGVVWVAFDKSSSLGQASKTIFVGVADPNPNVFWSRDGGATQAVPGAPFTIP 242
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 -----SVTFVLFDKSGSGKSGPTPIFGVCGSEPKG-IEVTEGDTTWNVLNPLNDLIP 239
QY 243 HKGVDPVNVHLYIATSNITGGYDGGSGGVWKFVSTSGTWTRISPVSTDTANDYEGYS 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 LRKGIH--DGILVTLUSNALNGATRGAVMKYVIADQKWVDYTPMKGD-----FGYCG 291
QY 303 LTIHQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTISYNNRSLRYLVIDISAEPW 362
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
292 IDVOE---NVIVSTLDRVPHDEIFISLNGGETWRPLEKANF-----DINKAPW 339
QY 363 LTFGVOPNPVPSKLGWDEAMADPFNSDRMLYGTATYATNLDLTKWDSGGQIHIAIP 422
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 IK-----HWISD-VKIDPFDMNRAIFTGYGVWVTELYKKSFEG-----MGK 384
QY 423 MYK-----GLEATVNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVETTC 476
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 PVKWIENKGLEETVVLQVLPPIGERPLLSATADMGFRHESLDTPPSSMY-KPLKWTSL 443
QY 477 SVDYAELNPSIIVRAGSPDPSSQPNDRHVAFTSDGKNNFQ-GSEPGVTTGG--TVAAS 533
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
444 GIAFAYSKFEVARVHTY---IYV---FLSYSDGGINNREIETVEGITDGGRLSLAYS 497
QY 534 ADGSRFVWAPGDPQPVYAVGEGNSWAASQGVPA---NAQIRSDRVNPKTFYAL---SNG 588
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
498 NDKGTLVWSPAN--HEVIVSSDKGSKWKKAISYPVPFNYFPASDPVNPFSKEYIFDWKNG 555
QY 589 TFRSTDDGVTQFQVAAAGLPSSGAGVGMFHA-----VPGKGGDLWLAAAS-SGLYHSTNGGS 643
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
556 DFLISKDGKSPMK-CAKLPSPDNWVWVSIYSPVLAPDREGDIWLALQWNLGRYSKDGGI 614
QY 644 SMSAITGVSAVNVGFKSAPSSYPAFVVGITIGVGTAYRSDDCGTTWVLINDDQHOY 703
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
615 TPERLGNVDIAVYIGFGAPKPGTDYPIYLVNGWVGVYGIENSTDGSKTWMRLNNDKHQF 674
QY 704 GNVGQAITGCHANLRVYIGTNGRGIYVDI 734
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
675 G-WIHYMIGDMNEFGRIFLTGREGRIIVEV 704
```

```
RESULT 7
Q912M3
ID Q912M3 PRELIMINARY; PRT; 2468 AA.
AC Q912M3;
DT 01-WAR-2001 (TREMBLrel. 16, Created)
DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA1874.
GN PA1874.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Stover C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004613; AAG05263.1; -.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR003880; Phosphopant_attach.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 5.9%; Score 238; DB 16; Length 2468;
Best Local Similarity 22.8%; Pred. No. 6.3e-05;
Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

QY 11 VAIGGGYDGVINEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWGMNNGY----- 65
Db 1084 VNLSGSSLSG---TAEPGSTVILT-----GNCNPITAEVTADSGNWTPTSTP 1130
QY 66 --NG-VVSTAADPINTNKVAAVGMVYNS-----WDPNDGAILR--SSDQATWQITPLP 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1131 IANGTVVYVAQDASGNSPPATVTVDSAPPAPVLPNSGVVISGTAEGAGATVTLT--- 1187
QY 116 FKLGGMPCRGMERLAVDPNDNINILFYGAPSGKGLWRSTDGATWSQNTNFPDVGTYIA 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1188 -DAGGNPIQGVTA-----GSGNWSFTP-----GTPLA 1214
QY 176 N-----PTDTIGYQS---DIQGVVWVAFDKSSSL--GOASKTIFVGVADPN-N 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1215 NGTVIVATATDPTGNTPOAATTVDAVAPPVDPIDPSNGTTISGTAEAGAKVILTDGNG 1274
QY 219 PV-FWSRGGATWQAVPGAP-----TG-----FIPKHGVDPVN 251
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1275 PIGETADSGNSWSEFTPGTPLANGTVVYVAQDAPAGNTGPGQSTTVDAVAPNTPVVNP 1334
QY 252 HVLVYATNTGGPY---DGSSGDVWKFVST-SGTW--TRISPVPS---TDTANDYFGYS 301
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1335 GNLLNCTAEPGSTVTLTDGNGNPIGTGTADSGNWSFTPGSQLPNGTVVYVNTASDAAGNT 1394
QY 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDG-----ATWTRIND 342
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1395 SLPATTTVDSSLPSIPQVDPNSGVISGTADAGNTIII--TDGNGNPICQVTDAGSGNWS 1452
QY 343 WT-----STPNRSRLRYLVIDISAEPLW--TFGVQPNPVPSPKLGWDEAMADPFNS 392
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1453 FTGPIPLPDGTVVNVVARSVNSVDSAPAVITVDGVAPAAPV-----IDPSNG 1499
QY 393 DRMLYGT---GATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISP-----PSCAPLI 445
```



```
Db 1500 TE-ISCTAEAGATVILT-----DGGCN-----PIGQATADGSGNWTFTPTPLANGTVIN 1548
QY 446 SALGDLGGFTH-----ADTAVPSTIF--TSPVFTTGTSDVYAEINPSIIVRAGSFDPDS 497
Db 1549 AVAQDPAGNTSGPASVTVDAIAPAPVNPNSGVVISGT-----AEAGATVILTDGNGNPI 1604
QY 498 SQPNDRHVAFTSDGGKNWF-----QSEPGGVV-----525
Db 1605 GQ-----VTADSGNWSFTPGTPLANGSVINALQAAGNNSPTSATVDSLAPAAPV 1657
QY 526 -----TCGTVAASADGSRFWAPDGPQPV--VTAVGFGNSMAASQGVPAANAQIRSDRVNP 579
Db 1658 IDPSNGSVIAGTAEGATVILTDGNGNPIGQVTAADSGN-WSFTPGTP-----1704
QY 580 KTFVALSNGTFYR-----STDGGV--TFQPVAAAGLP---SSGAVGVMFHAPVKGEGD 626
Db 1705 -----LSNGTVVNAVQAADAGANTSGPVSTTVDAVATPVIDPSNGVELSGTAEPGRVI 1759
QY 627 LWLAASSGLYHS--TNGSGSWSAITGV-----SSAVNVGFGKSAPGSSYPVAVFVGTIGGVV 681
Db 1760 LTDGNGNPIQTLADGSGNWSFTPGTPLANGTVVNAVQAQDPAGNTSGPASTTVDTVAPAT 1819
QY 682 GAYRSDCGTTWVLINDDQHQYGNWGOATIGDHANLRRVYIGTNGRGVYVGDIGGAPSG 740
Db 1820 -----PVINPSN-----GSVITGT-AEVCAKAVILTDGNGNPIGETTADGSG 1859

RESULT 8
Q91948 PRELIMINARY; PRT; 1800 AA.
AC Q91948;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-BINDING OUTER MEMBRANE-LIKE PROTEIN (FRAGMENT).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225830; PubMed=10762233;
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RT "Genetic analysis of functions involved in adhesion of pseudomonas putida to seeds."
RL J. Bacteriol. 182:2363-2369(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF182518; AAF69021.1;
DR InterPro; IPR003644; Calx_beta.
DR InterPro; IPR001343; Hemolysin_Ca_bind.
DR Pfam; PF002035; VWFA.
DR PRINTS; PR00353; hemolysinCbind; 3.
DR SMART; SM00313; CABNDNGRPT.
DR SMART; SM00237; Calx_beta.1.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 2.
DR PROSITE; PS50234; VWFA; 1.
FT NON_TER 1
SQ SEQUENCE 1800 AA; 185196 MW; 409BFC017552F96E CRC64;
```

```
Query Match 5.3%; Score 214; DB 2; Length 1800;
Best Local Similarity 22.0%; Pred. No. 0.00093;
Matches 186; Conservative 107; Mismatches 319; Indels 232; Gaps 41;
QY 12 AIGGGGFDGIVNECAPGLLYVRTDIGMYRWDAAANGRWIPLLDVWGNMNCYGVVSI 71
Db 854 AAGNLOFFNGTAVNVAVGVVQSADI-----TAKNLKFPALNOSGADNYGGNGVGNQ 907
```

```
QY 72 AADPINTNKVMAAVGYTN-SWDPNDGAILRSSDQATWQITPLPKLGNMPPGRGMGER 130
Db 908 KAD-----YAQFKFKPNDGTNL-----GSEVTMK 931
QY 131 LAVDPNNDN-ILYFGAP--SGKGL----WRSTDGATWSQMTNFPDVGTYIAN-----P 177
Db 932 VDISPVADRPKTLFSGSADIESKGLTKREVMTSLKGLGTGGNGITGDLKTVFANSANS 991
QY 178 TDTTGYQSIDIQGVVVAEDKSSSSLSGOAKT-----IFVGADPNPNPFWSRD 225
Db 992 SITTNVQSD-----GSVTAGTKTSLGLIYLEAGKVVYTFGLADDS---FWVTI 1037
QY 226 GG-----ATWQAVPGAPTGFIPHKGVFDPVNVHLYIATSNMTGGPYDGGSDVMKFSVTS 280
Db 1038 GKGTVVATWAGGGVSGTFTPTNTSGYPIE--VYHANQSGPGSYD-----LNIQVGS 1089
QY 281 TWTRISP-----VPSTDANDYFGYSGL-TIDRQHPNTIMVATOISWMPDITIFIRSTDG 333
Db 1090 AVTDLSSSNVQYQNTYEMANAGLVSDLHTVNGQ-----SYDGYKLNKPEG 1138
QY 334 G-----ATWTRIMDWTSPNBSLR-----VYLDISABEPLTFGVQPNPPVSPKLGWM 381
Db 1139 GSVKLVGIETALDTDGSESLNVTLSGIPKGTVLSDGAGHTVTVGT-----APVDVTGW 1193
QY 382 DEAMADIP-----FNSDRMLYGT-----GATLYAT-----NDLTK 411
Db 1194 LSSLTLTPPAYKGSFDITVTSTATESLGSAITTCNIPVTYVYATYKASVGTSGNDTLT 1253
QY 412 WDSGGQIHIAPMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPTITSPV 471
Db 1254 GSEGNDFIVAD-VSGLVNVQGNKYN-----NIAFMVDSGSGMSDKSIADAKTQLASVFNTLK 1308
QY 472 FTTGTS-----VDY-AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKNWFOGS-- 519
Db 1309 ASLGSSTGTVNIFLVDFDTQVKNKNAVNLA--DPDALSCLKQVLNMSVGG--YGGGTN 1364
QY 520 -EPGGVTT-----GGTVAASADGSRFWAPDGPQPVYVAVFGNSMAASQGVPAQAIRS 574
Db 1365 YEDARFTTSNFFNSTWATSNKGAENLYFTITDGKPYQO---SNSTNPISLWKNKSLD 1420
QY 575 DRVNPK-----TFYALSNGTFYRSTDDGGVTFQPVAAGLPSSCAVGMHVAFCCKEGDLW 628
Db 1421 DVVNVNMYKMGDTFSAWADATH-----KVEISSSGVVKVLTAY-TENRRGELV 1466
QY 629 LAASS--GLYHSTNGGS-SWSAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGVTAIR 685
Db 1467 LUSTKTVGLTHAQDGTTFESSLDTGTYADYNNVYISAAGST--ESFAV--LGGTNGLSK 1522
QY 686 SDDCGTTWVLINDDQHQYGNWGOAITG-DHANLRRYIIG-----TNGRGIV 730
Db 1523 VQAIGLNSDVTLNDLKPYSAGKPKQTNIDPSDLAKAILGHSEATVPGADTIDSGNGNDII 1582
QY 731 YGDI 734
Db 1583 FGDL 1586

RESULT 9
Q92008 PRELIMINARY; PRT; 2174 AA.
AC Q92008;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL GLYCINE-RICH PROTEIN SMB21548.
GN SMB21548.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
```


RC	STRAIN=1021;
RX	MEDLINE=21396508; PubMed=11481431;
RA	Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
RB	Vorhoeher F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA	Golding B., Puehler A.;
RT	"The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT	fixing endosymbiont Sinorhizobium meliloti.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9899-9894 (2001).
KW	EMBL; AL603645; CAC49389.1;
DW	Plasmid; Hypothetical protein; Complete proteome.
SQ	SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match	5.0%;	Score 202.5;	DB 16;	Length 2174;
Best Local Similarity	22.7%;	Pred. No. 0.0052;		
Matches 176;	Conservative 90;	Mismatches 312;	Indels 197;	Gaps

QY	12	AIGGGFVDGIVENEGAPGILLYRI-DIGGYRYRDAANGRWIPLLDWGNWNWCYNQVVS	70
Db	976	SIIGGG-----GNGGVGSVKSEIKGFNLT-----NVGVGGSGS	1011
QY	71	IAADPINTNKVAAGMYTNSMDPDNDGAILRSSDQA-----TWQTLPPEK	117
Db	1012	GGASGNN---AIVGL-----DSCTHLOTSGSARGEVIVQSIGGGGTSGCASVGLS	1059
QY	118	LGNMFGRMGERKLAVDPNNNILFYGPSKGWLWRSTDQSQTWSQMTNFPDVGYTIANP	177
Db	1060	ASASLPGGG-EAAAEAEESGSGAFSAVSQSVSGRTGGSGGS-SGTVNVTACT-----	1112
QY	178	TDTTGYSODIOGVVVWFDPKS---SSSLGOASKTFIVCVADPNPNPFWSRDG-----	226
Db	1113	--ISTGAADAGVLQAISIGGGGLGSGVASGGSEPDLDDDEGSECAGEGNGDGHGYY	1170
QY	227	-----GATWAQVPCAPTGFPHKGVFPVNHVLIATS-----NTGGPYDCSSG	270
Db	1171	CFGVSVGAT---IDCGTGTAANGNAVTLFHAGHIATAGDWDAGIVAQSIGG--GGAG	1225
QY	271	DNKFSVTSWTWTRISPVPSTDTANDFYGYSGLTIDROHPNTIMVATOISMPDITIFRS	330
Db	1226	GTSTAQSQATANIITVVGSGGAGGNGGAVGITFDNHNHSISTAGYSAY---GVLLQS	1282
QY	331	TDCGATWTRLDWITSYPNBSLRVLD-----ISAEPWLTFCVQNPNPVPS	375
Db	1283	IGGGGG-----QGSDGDEAAGRITVGGGFGSGGAGGSGGMVTKAGINLTSG-----	1332
QY	376	PKLGWMDEAMAI--DPFNDRMLYGTGATLYA-----TNDLTKWDSGGQIHIAPMVKGL	427
Db	1333	-----DDAHGIVAQSIGGGGVGAGSSTAEREHSHTIDLTVVGGSG-----VGGSGGE	1382
QY	428	EETAVNDLISPPSGALISAL-----GDLGGFTHADVTAVPSTIFTSFVTTGTSDVY	480
Db	1383	VDSLVSQTTLS-TSGARALGLVAQSIGGGGGIGGAGEAD--SIASLVVGG---SGGGTIDG	1436
QY	481	AELNPSIIVRACSFDPSPQNDRH--VAFSTDGKN-----WFOQSEPGGVTTGG	538
Db	1437	GAVTYDLTQSQSSI---TTOGIAAHLVAQSIGGGGGVGGGAASGAPLFTGNSPGSYGDGG	1493
QY	529	TVAASADGSRFV--WAPCDGPQPVYAAGVGENSWAASQGVPAQAQRSDRVNPKTFYAL	585
Db	1494	DVAVTAGDSIIFRGDYAFGLAQSIGGGGGFEGGN-ATSAFTGSGNGLUSSDG-----	1543
QY	586	SNGETFYRSTDGGVTFOPVA-----AGLPSSGAVGYMFH-AVPGRKEG----LWL	629
Db	1544	KSGNVTVSLDAGRTIQASCCKDSIGIFAOSDAGTDNNGTIDTVVNGTYTGGSDNGAGIWW	1603
QY	630	AASGLYHSTNGGSSWASITGVSSAVNWYFKSKAPSSYPAVFVVGTI-GEVGTGA	683
Db	1604	SAGKDNNVTVNSGGNYSAASGV--AVQYTAGNNSPEDSTLVVNNAGTISGSVKGA	1656

RESULT	10
ID	P73139
PRELIMINARY:	PRT: 3972 AA.
ID	P73139

AC	P73139;
DT	01-FEB-1997 (TrEMBLrel. 02, Created)
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	HYPOPHETICAL 418.3 KDA PROTEIN.
GN	SUR1028.
OS	Synechocystis sp. (strain PCC 6803).
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX	NBI_TaxID=1148;
EN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97061201; PubMed=8905231;
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA	Miyajima T., Hirose M., Sugilura M., Sasamoto S., Kimura T.,
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA	Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
EA	Tatata S.;
RT	"Sequence analysis of the genome of the unicellular cyanobacterium
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT	entire genome and assignment of potential protein-coding regions.";
RL	DNA Res. 3:109-136(1996)
DR	EMBL; D90904; BAA17165.1; -.
DR	HSP: P30617; LJPC.
DR	InterPro; IPR001480; B_lectin.
DR	InterPro; IPR003644; Calx_beta.
DR	InterPro; IPR000413; Integrin_alpha.
DR	InterPro; IPR000215; Serpin.
DR	Pfam; PF01453; Agglutinin; 1.
DR	Pfam; PF01839; FG-GAP; 8.
DR	PRINTS; PR01185; INTEGRINA.
DR	SMART; SM00108; B_lectin; 1.
DR	SMART; SM00237; Calx_beta; 1.
DR	SMART; SM00191; Int_alpha; 9.
DR	PROSITE; PS00284; SERPIN; UNKNOWN 1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 3972 AA, 418281 MW, B9FE49P274FBEABD CRC64;

Query Match	4.7%;	Score 189;	DB 16;	Length 3972;
Best Local Similarity	19.3%;	Pred. No. 0.063;		
Matches 166:	Conservative 93;	Mismatches 276;	Indels 326;	Gaps

Qy	13	IGGGGFVDGIVFNAGPAGILYVTRDIGMYRWDANG-----RWPLLDW-	57
	:	: : : : : : : : : : : :	:
Db	2458	VNGDGFDADVSGGSPAGGVLIIGNSTKDL--DAALGTDLLIISVENAQVFALGDGF	2515
Qy	58	-----VGWNWCYNG-----VVSIADPINTKNVVAVGMYTNSWD-	93
	:	: : : : : : : : : : : :	:
Db	2516	GDGLADEGVIDDDGNFFLVLSPELGSQSLVIDSTLPNLNFAWGVDENGVDYDF	2575
Qy	94	----PNDGAILRSDDCATWQIPFLFKLGNNP-----GRGMGERLAVDNP-	136
	:	: : : : : : : : : : : :	:
Db	2576	VLCQPNSTIAVYGNGATLTDSPLTF--GNFPPLPSFTGIDLNGNIKEIVAGQPNLN	2633
Qy	137	-----NDNILYFGAPSGKLWRST-----DSG-ATWSQMTNFPDVGYTIANP	177
	:	: : : : : : : : : : : :	:
Db	2634	PVPNIGGFGGGLYFYEAGNAVLOFTVPPNPASVTEARSSGLSWQI-SFPNQYAAGVP	2692
Qy	178	TDTT--CY-----QSIDIQVWV-----VAFDKSS-----	200
	:	: : : : : : : : : : : :	:
Db	2693	SFATLDGWLYQAIFYGNERISTKDSYIIQRSDRGVSLENLTQQVPLDSNGTFLDKNLP	2752
Qy	201	-SLGOASKTIFVGVAOPNNPVFWSRDG-----GATWQAVP-----GAPT--GFIPH	244
	:	: : : : : : : : : : : :	:
Db	2753	PSITAYNGTLVLGTTADNGOV-WVAGEVVTNANSGLILNAVPIQNASNGTFLVAFNDEL	2811
Qy	245	GVF---DPNVHVLVIATSNTPGYDGGSGVKFSVTSGWTRISPSTPDTDANDYFGY	300
	:	: : : : : : : : : : : :	:
Db	2812	YVFFVKDASDNILYSSSNPG-----SSSG--WDGSTVLFVSDYNQATN-----FPL	2858
Qy	301	SGLTIDRHQPNTIMVATQISWWPDTIIFRSTDGGATWTRI-----WD-----	342
	:	: : : : : : : : : : : :	:
Db	2859	SATVPGILDGTTLAVA-----FRSNSNPATWGLLSSDVTNWOGSAELTOVDA	2907

RESULT 10
P73139
ID P73139

QY 343 -----WTSYPNRSLRYVL-----DISAEPWLTFGVQPNPPVPSKL 378
Db 2908 NSQVSLFWVDGTYLFTFSSTASASTDGLNWGDITLIPW-----DDGNL 2956
QY 379 GWDE-----AMADPFNSDRMLYGTGATLYATNDLTKWDSGGQIH----- 419
Db 2957 GGVASILLFNSQFSLSLAQSNESLLFAFSLPEPNQASRW--GEQVRDIDGDFNGDGIAD 3014
QY 420 ---TAPVWKLEETAVNDLISPPSGAPLISALGDL---CGFTHADYVAVPSTIFTSVPF 472
Db 3015 LAVLAPGRNLLQPILDY-----FAINNLGGVFYIYGEESGISVNDPPDVVLAAPDL 3067
QY 473 TTGTSVDAELNPSIIRAGSFDP--SSQP-----NDRHVAFTSDGKWN-- 515
Db 3068 PQETIFELLEITPGDVGNGDFDGLLLSAPLTPVIAQGFDPVNGDQGVSWVFGTGWGT 3127
QY 516 -FOGSEP-----GGVYTG-----GTVAASADG--SRF-VW 541
Db 3128 EYTANSPFGLNLANNQTNNSQNPENYGFVTTGLPRSQAGISISGGADVNGDGFSDFALG 3187
QY 542 APGD-----PGQPVYAVGFGNSWAASQG---VPANNAQIRSDR 576
Db 3188 APGPNLNSVLFSGDFTNQVNLGTIGTDVMLGSPTEIFVAGQDDQIYTNNGVDTVY 3247
QY 577 VNPRTFYALSNGTFYRSTDDG 597
Db 3248 AGPNDFTVTDTNFRRLDGG 3268
RESULT 11
P74440
ID P74440 PRELIMINARY; PRT; 4199 AA.
AC P74440:
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEtical 442.4 KDA PROTEIN.
GN SLR0408.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiyama M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res 3:109-136(1996).
DR EMBL; D90915; BAA18541.1; .
DR InterPro; IPR002860; BNR.
DR InterPro; IPR003644; Calx_beta.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002989; Mycobact_pentapep.
DR Pfam; PF02012; BNR; 3.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF01469; Pentapeptide_2; 1.
DR SMART; SM00237; Calx_beta; 1.
DR SMART; SM00191; Int_alpha; 8.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4199 AA; 442441 MW; 1DB4BC556D2C1B5C CRC64;

Query Match 4.6%; Score 186.5; DB 16; Length 4199;
Best Local Similarity 19.2%; Pred. No. 0.093;
Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

QY 9 SNVAIGGGFVDGIV-----FNEGAPGILYVFTDGGMYRWDAANGRWIPLL 55

Db 2361 ADVASDNGFVDGNLIGNPPTTPTTTSQYIDTTPAILINGSNLYAYKFGGNNQ---IY 2317
QY 56 DWYGNWN-WYNGVNVSTAADPINTKWN-AVGMWYNSW-----DPNDGA-ILRSSDQGA 107
Db 2318 FTVSTNNGSWNSEVOL---PQSAQTIFPPAIAFFNNVLAYVVDGNNGLNIITSDQGG 2374
QY 108 TWQITPLPFLKLGNNMPG-----RG-MGERLAVDPNDNLIYFGAPSGKGLNRSTOSGA 159
Db 2375 TWN---APLALGGTSSPTPLFVYQGTLSLLFRANNSTSTVQLFYLNSSNEWIYANEIGS 2431
QY 160 TWSQMT-----NFPDVGTVIANPTDITGTQSDIQGVVWVAF----- 195
Db 2432 NQTAISASATVLGDTLYLVKGGTRNTPSTLDYITSTN-----ADLSANDWSSIPIGP 2486
QY 196 --DKSSSLGQASKTFVGVADPNPNV-FWSRGGATQAVPGAPTGF---PHKGVFDP 249
Db 2487 VSSQGGPSLNDGNTNLVLSYLDSSNQLNFVSSGNGINWSS-PQVITNNISQSPPAIAF-- 2543
QY 250 VNHVLYTATNTGGPYDCSSGDVWKFSVTGWTTRISVPST-----DTAN 295
Db 2544 ANNELYLS-----YPCQGS-QELNVT-----PPLFTGSLGNGSLVRLGDVNG 2589
QY 296 DYFG--YSLGUTIDRQHPNTIM-----VATQISWNPDTIIFRSTGGATWTRIDWTSY 346
Db 2590 DGFADVFSGGT---NAGAIIFGNSTKDLTLTASGSDLVI-----SV 2628
QY 347 PNRSRLRYVL---DISAEPWLTFGV-----QPNPPVPSPKL 378
Db 2629 PNATLRDVISVDFNGDGIKDLGVLDGNGNFYVVLGNTSLGLDKTLISITSSSPVVINQV 2688
QY 379 GWMDEAMADPFNSDR---MLYGT-----GATLYATNDLTKWD----- 413
Db 2689 GGVTKSMAIGYNGDGYDDVLLMGDNGNOVANGNSTGVLSNFTNIDYPTQTATGVLDN 2748
QY 414 -----SGQTHIAPMWKLEETAVNDLISPPSGAPLISALGDLGGFTHAD 458
Db 2749 SDGIPEIAIGSDERKIAQISTSGFSLLPTTSSVINTLAAANQLENIGDFNGDGIAD 2808
QY 459 VTAVPSTIFS-----PVFTGTSTVDYAEIN-----PSIIVRAGSFDP 497
Db 2809 LAVIASNYAAAGEPNLNPYLRPNQNGVFIYFNGSNGLSNTAQPVDVILAAPFNPS 2868
QY 498 SQ-----PNDRHVAFTSDGKKNW----- 515
Db 2869 QISTYQLSRIAQAGDVNGDGFDDLLISPYTVDAENNGGVFVFGDDNNQPFDLGQ 2928
QY 516 -----FQGSPEPGVTTGGTVAASA-----DGSRF-----VWAPG----- 544
Db 2929 LRANQSGSNPRGFAIDGSPNSQAGIALNGGGDINGDGFADFIIGAPGENNLQYNOQIVF 2988
QY 545 -----DPGQPVVYAVG-----FGNSWAASQGVPAANAQIRSDRVN---PKTFYAL 585
Db 2989 IENGELSDDDKYSILYLDGNQTIOMGGGDWQANQ-VWTN-QVATWNNNSRPRPEAVIGQ 3046
QY 586 SNGTFYRSTDDGVTFO-----PVAAGLPSSGAGVGMFHAVPGK 623
Db 3047 SNGDIWYIPGNGNQWQSGWGLPAEINELAVNWNSTSGNPQIIAGLGGKG--GIEVY----- 3099
QY 624 EGDWLAASSGLYHSTNGGSSW-SATGVSSAVNVGFGKSAKPGSSYPVAVTVGTIGVTG 682
Db 3100 NGSTWV--NNGPYQ---GDGWRSAITQMA---VOMGED--GS--PSQIVVGLADGAVI 3145
QY 683 AYRSDDDGTTWVLINDDQHOYGNWGOAITG-----DHANLRVYIGTN 725
Db 3146 YYNTQ---SGWRTIN-----NFGKSVTOLSVQWQEPASNPVIVGLDNSEV-QYTOGSN 3194
QY 726 GRGIVYGDIG 735
Db 3195 GVWTQFHDG 3204
RESULT 12


```
Q9N1P0
ID Q9N1P0 PRELIMINARY; PRT; 818 AA.
AC Q9N1P0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE SUBMAXILLARY MUCIN (FRAGMENT).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BREED ANGUS; PubMed=10759843;
RX MEDLINE=20223253;
RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts.";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL; AF178428; AAF67279.1; -.
FT NON_TER 1
FT NON_TER 818
SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;

Query Match 4.6%; Score 185.5; DB 6; Length 818;
Best Local Similarity 25.9%; Pred. No. 0.014;
Matches 159; Conservative 55; Mismatches 249; Indels 151; Gaps 34;

QY 120 GNMPPGRMGERLAVDNNNNILYFGA-----PSGKLMRSTDSGTATWSQMTNFPDVGTY 173
D 182 GSSPGRSRATAVSGESQPTVALSGATGTSAGPSGT---RSTSSAIP----- 225
QY 174 IANPTDTTGQSDIOGVVWVAPD-KSSSLGQASK--TIFVGADPNPNFVNSRD---GG 227
D 226 -ATPGSTTGRAA---GAGTAVDSQQTASLPAAARTALPGTSPAGTSESRSVPGG 281
QY 228 ATWQAVPGA---PT-----GFPHKGVDPVNVHVL--YIATNTG---G 263
D 282 SETTQPGAGSEPTLSPGVTRTALRGSETRVPSGTGVSLPGSTOGGSAATGSGAGSG 341
QY 264 PYDGGSGDVWKFVSGTWRISPVSTDTANDYFYGSLTIDRQHPNTIMVATQISWNP 323
D 342 PTAPVSGET-RTSVISGTVNVPVSCAPVTPGSS--AGSSGA-----PGT-----GGP 384
QY 324 DTIIFRSTDGGATWTRIDWNTSYPNRSLRYVLDSIAE-PWLTFGVQPNPVPSPKLGWMD 382
D 385 GSETASPLSAAATGATGSGTSLPPSG---APVTPEPPLISTGASAGPPASSESTVTL 440
QY 383 EAMAIIDPFNSDRMLYGTGATPLATNDLTWKDSGGQIHAPMVKGLEETAVNDLISPPSG- 441
D 441 GATGTDVLR-----GTSLPV-----SGGAVTPAPSPGSSATA-----GPCVGS 480
QY 442 APLISALDGLGGTHADVTAAP-STIFTSVFTTGTSDVYAEINPSIIVRAGSDFSSQP 500
D 481 ATTQVQSGATGADVLRSGTSLPVSGVAVSPGSSPG-----RSGTAVSS-- 524
QY 501 NDRHVFSTDGGKNWFGSPGGVTTGGT--VAASADSGSREFVWA--PGDPCQPVVYAVGFG 557
D 525 -----OGSQTVALSGATGTSVGPSTREFSSAIPATPGSTTGRAAGAG 568
QY 558 NSWAAASO---GVPANAIQRDRVNPRTFYALSNGTFFVRST-DGG--VTFQPVAAAGLPSSG 611
D 569 TPGVDSQQTASLPAA--RTALPGTSPAGTSESRSVPGGSETTQPGAGSEPTL 626
QY 612 AVGV-MFHAVPGEGDLWLAAASHSTNGGSSWSAITGVSAVNVGFGKSPGSSYPA 670
D 627 SPGVTRTALRGSETRVPSGTGSLPGSTOGG---SAATGSSGA---GSGTAPVSGETR 680
QY 671 VFVV-GTIGGVTGA 683
```

```
Db 681 TSVISGTVNVPVSGA 694

RESULT 13
Q92411 PRELIMINARY; PRT; 997 AA.
AC Q92411;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CELLULOSE PRECURSOR (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00598; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1
FT SIGNAL 34
FT SIGNAL 34
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 4.6%; Score 185.5; DB 2; Length 997;
Best Local Similarity 21.1%; Pred. No. 0.018;
Matches 165; Conservative 85; Mismatches 248; Indels 285; Gaps 43;

QY 79 NKVMAVGMVNTS---WDPNDGAILRSSDQOGATWQITPLPFKLGNNMPGRGMGERLAVDP 135
D 147 NELWQGVGAGTNDHAWMPAE-----VQMNRPSFKIDASCPCGSDLAETAAL 195
QY 136 NDNILYFGAPSGKGLWRSTDSGTWS-----OMTNEPDV--GTYIANPTDT----- 180
D 196 AASIVF-----ADSDPVYSAKLLQHAKELYNFADTYRGKYTCIIDAFAFYN 243
QY 181 --TGYQSDIQ-GYVWV-----AFDKSSSLGQASKTIFVGVADPNPNPVFV 222
D 244 SWTGYEDELAWGAWLYLATNDNAYLSKALSAADRWSTSGSA-----NWPYTW 292
QY 223 SRDGGATWQAVPGAPTGFPHKGVDPVNVHLYIATNTGPGDYDSSGDVWKFVSIVSGTW 282
D 293 TQG-----WDS-----KHG---AQILLARITSNLNP-----EATKF----- 322
QY 283 TRISPVSTDTANDYF--GYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGGATWTRI 340
D 323 -----IQSTERNLDYTWVTGNGGRV-----KTPPGGLAWLDQ 354
QY 341 W-----DWTSPNRSRLRYVLDSIAEFPWLTFFGVQPNPVPSPKLG---- 379
D 355 WGSRLRYAANAATFSFVYSDNWVSDPVKRSRY--QNFAISQINYILGDNPROSSVVGYGQN 412
```



```
QY 380 -----WMDEAMAIIDPNSDRMLYGT-----GATLYATNDL-----TKW 412
Db 413 SPQPHRTAHSWMNEDI--PANHRHILYGAMVGPNASDQYDDIGDYSVNEVATDY 470
QY 413 DSG-----GOIH-----TAPMVKLE-----ETAVNDLISPPSG 441
Db 471 NAGFTGALAKMNLGYGNHQPANFPAPEVKGYEYFVEAAVRSSGSGNYTEIRALLNRS 530
QY 442 APLISALDGLGGFTH---ADVTAVSTIFTSVFTGTSVDYAEINPSIIVRAG----- 492
Db 531 WP--ARMGDOLSKYFLDLSEVTAAGRTV--SDVQTVTSSEGAATVQPVVDAAKRIYA 586
QY 493 -----SDPPSSQPNDR-HVAFSTDGKKN-WFQSEPG--GVTTGG----- 528
Db 587 ITANFSTKIYPGEGNYRKEVQFRITGPOGANPANDPSYQNLTTGPNVKSNIYPYDA 646
QY 529 --TVAASADGSRVFWAPGDPQPVVAVGFGNS-----WAAASGVFANAQIRSD-RVNP 580
Db 647 GKVVSQEPGVTFVAVPAAPAG--VOAVA--GNSQVALNWSASAGAVSYTVKRAEVS 703
QY 581 TFVALS-NGFYRST--DGGVTFOPVAAGL-----PSSGAVGVMFHAVPGKEGDL 631
Db 704 TTVAAVGNGLTNTYTGTLNGKTYIYVTVAVNSAGESPASVQVSCMPQAAATVPCAV 763
QY 632 SSGLYHSTNGGSSWASITGVSSAVNVGFGKSPGSSYPVAVFV-----VGTIGV 682
Db 764 TAG---NNQNLWSATAAGSSYT---VQRAVAGTYTVDATGLAVLNYNDTTALNGT 817
QY 683 AYR 685
Db 818 SYR 820

RESULT 14
Q9HLO9 PRELIMINARY; PRT; 998 AA.
AC Q9HLO9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THERMOPHILIN PRECURSOR RELATED PROTEIN.
GN TA0167.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmales;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Wewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445063; CAC11313.1; -.
DR MEROPS; A05.0PW; -.
KW Complete proteome.
SQ SEQUENCE 998 AA; 108669 MW; EC2E9E92E382853 CRC64;
```

```
Query Match 4.6%; Score 184.5; DB 17; Length 998;
Best Local Similarity 20.1%; Pred. No. 0.02;
Matches 173; Conservative 113; Mismatches 285; Indels 289; Gaps 47;

QY 9 SNVAIGGGGFGDGVFNEGAPGI-----LYVRTD-IGG----MYRWAANGRWIP- 53
Db 188 ANALHGNGKLVGVYFYDGPTEFNISFPLKLYLNSSLVGGNDAYFNYSIENGKITKS 247
QY 54 -LLDNVGNWNGYGVVSTAADPINTNKVWAAVGMVTNSWDPNDAILRSSDQATWQIT 112
Db 248 GSYDAVTFNSSG-----KASSPARFLISGYSPPAGLLYDAELAITG--- 289
```

```
QY 113 PLPFLKGGNNPGRGMRGERLAVDPNNDNLYFGAPSGKGLWRST-----DSGATW 161
Db 290 -----PGGG-----SNVNYGICNGTGLYLYNSTSKAFNTVRSAYDAGVDT 330
QY 162 SQTNPFDVGTYIANPTDTGTGYQSDIQGVVWVAFDKSSSLGQASKTIFGVGADPNPVF 221
Db 331 GETSYGVDVGWY--GNMADLQSGPSLIYG-LW-----NVSSGIETI-SGTAEPSQYIF 379
QY 222 WS-----ROGGATWQAVPGAPTGFPHKGVDPVNHVL-----YIATSNTPGPDGSS 269
Db 380 VSNKGFNNSTASWAPYSGAHFRKLPKGSYSLA--VLRNLYAPEYINDAGTSPSLNGTS 437
QY 270 --GDVWK-----FSVTSGTWTTRISPV-----PSTDANDYF--GYSG 303
Db 438 KTGDIVAPVIENYSDALYYSRSGNSQISPFVIFSGQTVSDPFLGVLNDLYFPVFTGL 497
QY 304 T-----IDRQHPNTIMVATQISWNPDTIIFRSTGGGATWT-RIMDWTSYPNRSRYVL 355
Db 498 AIINTSYHIDVENASSEFEITYGSQY--DSLASHGLGNTNMGMILYGTSS--NVSITGGL 553
QY 356 DISAEPWLTFGVQPNPPVSPKLGWDEAM-----AIDPFNSDRMLYGTGATL 403
Db 554 NIT-----GWPEPMKGFVAVANLWVWNSNDLIADNVFY-----SL 589
QY 404 YATNDLTKWDGGQIHIAPMVKGLEETAVND-----LISPPSGAPL-----ISALGD 450
Db 590 HASSS---DLOPPYSSVLLIYGMNSTRANTVWGNVFPDQYVSENGNATVGLFVDSSGN 645
QY 451 L---GGFTHADVTAVPS-TIFTSPTVTTGTSVDYAEINP---SIIVRAGSFDSSQPNDR 503
Db 646 LIYNNAFTQYDIAFSPDFNIYTDNATYHDNWNISSEVPIYNTINVGYSL-----NGS 699
QY 504 HVAFTSDGKKNWFGSGEPGGVTTGGTVAASADGSRFWAP-----GDPQPVV--- 551
Db 700 IVGGYEGGNYW-----GSPITPWNESGYIDSDYHPLVYFL 739
QY 552 YAVFGNSWAAASQGVFANA--QIRSDRVNPKTFYALNSGTFFYRSTD---GGVTFQPVAA 606
Db 740 YELSF-----AIHGLPAGTSWELVADGV-----YVVGSGPYINITEFNFGNTTY---AVF 786
QY 607 LP-----SSG-----AVGVMEHAVPGKEGDLW-LAASGLYHSTN 640
Db 787 LPPGYISNAGSAYAYATDMSTVNVYARHWYTGIVFATNYVKNETWYLDIAGLSYSTS 846
QY 641 GGSWSAITGVSSAVNVGFGKSPGSSYPVAVFVVGITGGVT-----GAYR----- 685
Db 847 QDLTVSLPNGTYS-----YSVFAPPNNYTAVSGYVTVNGVSEIVNITFPQLRYVIRET 901
QY 686 SDDCGTTWVLINDDQHOYGN 705
Db 902 GLPAGSSWYVVDGKAYYSN 921

RESULT 15
Q9RK65 PRELIMINARY; PRT; 848 AA.
AC Q9RK65;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SCF11.25.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:50 ; Search time 101.03 Seconds
(without alignments)
97.848 Million cell updates/sec

Title: US-09-917-376-4
 Perfect score: 470
 Sequence: 1 VSGGVKVOYKNNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

1:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/cgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231.5	49.3	782	AAR15625	Cellulase AE-1. A
2	195.5	41.6	616	AAV13494	Truncated cellulase
3	194.5	41.4	1751	AAV13493	Truncated cellulase
4	191.5	40.7	1426	AAV13492	Truncated cellulase
5	186	39.6	499	AAR42122	NK-1 cellulase. B
6	172	36.6	700	AAR13227	Novel endoglucanase
7	165.5	35.2	551	AAW18790	Corrected Bacillus
8	165	35.1	167	AAR95080	Cellulose binding
9	165	35.1	476	AAV34123	A mannanase-linker
10	165	35.1	493	AAV28850	Pectate lyase-link
11	165	35.1	493	AAV32318	Pectate lyase CBD

12	165	35.1	531	18	AAW15238	Scaffoldin protein
13	165	35.1	1853	19	AAW43108	C. thermocellum C
14	155.5	33.1	1352	22	AAAG63962	Amino acid sequenc
15	150.5	32.0	1350	22	AAAG63963	Amino acid sequenc
16	120.5	25.6	531	16	AAW01503	60 kD endoglucanas
17	120.5	25.6	532	12	AAW13229	Endoglucanase enco
18	114	24.3	162	15	AAW63634	Cellulose binding
19	114	24.3	162	20	AAW90077	C. cellulovorans C
20	114	24.3	163	22	AAE05745	Clostridium cellu
21	114	24.3	256	22	AAE05745	Clostridium cellu
22	114	24.3	328	22	AAE05749	Chimeric S peptide
23	114	24.3	341	22	AAE05747	Clostridium cellu
24	114	24.3	428	22	AAE05748	Clostridium cellu
25	113	24.0	190	22	AAE05746	Clostridium cellu
26	113	24.0	382	20	AAE39952	Gaussia luciferase
27	112	23.8	154	20	AAW90081	C. cellulovorans C
28	112	23.8	156	20	AAW90080	C. cellulovorans C
29	75.5	16.1	986	21	AAW40440	Human brain-derive
30	75.5	16.1	1346	22	AAU04567	Human G-protein co
31	75.5	16.1	1346	21	AAU05581	Human G-protein co
32	75.5	16.1	1371	22	AAU04570	Human G-protein co
33	75	16.0	1723	22	ABG24680	Novel human diagno
34	72.5	15.4	69	22	AAU17759	Novel human respir
35	70.5	15.0	2955	19	AAW56450	Fragment F1029 of
36	69	14.7	2970	19	AAW56445	Fragment HGU1916 o
37	68	14.5	2873	19	AAW56441	Fragment HGU606 of
38	67.5	14.4	857	17	AAW5287	Nasturtium xyloglu
39	67	14.3	236	15	AAW45442	Sequence of the si
40	66	14.0	237	17	AAW94020	Anti-erbB2 scFv.
41	66	14.0	237	18	AAW15185	Single-chain anti-
42	65.5	13.9	1566	20	AAW06309	Human p53 sequenc
43	64	13.6	117	20	AAW86130	Protein sequence o
44	64	13.6	143	20	AAW50155	Murine monoclonal
45	64	13.6	254	22	AAW64503	Drosophila melanog

ALIGNMENTS

RESULT 1	
AA15625	
ID	AA15625 standard; Protein; 782 AA.
XX	
AC	AA15625;
XX	
DT	17-MAR-1992 (first entry)
XX	
DE	Cellulase AE-1.
XX	
KW	Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
XX	
OS	Aeromonas strain no. 212.
XX	
PN	JP03251174-A.
XX	
PD	08-NOV-1991.
XX	
PF	28-FEB-1990; 90JP-0045465.
XX	
PR	28-FEB-1990; 90JP-0045465.
XX	
PA	(OJIP) OJI PAPER KK.
XX	
WI	1991-373412/51.
DR	N-PSDB; AAQ15176.
XX	
PT	Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs
XX	
PT	of opt. pH when carboxymethylcellulose is used as substrate.
XX	
PS	Claim 2; Fig 3; 8pp; Japanese.
XX	
CC	The sequence was deduced from the gene which was sequenced from
CC	plasmid, pAEC 1 prepd. by ligating chromosomal DNA contg. the
CC	gene.

CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
CC methylcellulose is the substrate. The N-terminal sequence: GIHADP-
CC has been confirmed by Edman degradation. The gene can be used to
CC produce recombinant enzyme which is used for the effective utilis-
CC ation of biomass resources and the mfr. of pharmaceuticals and
CC foodstuffs, and also for the detergent and deinking of waste paper.
XX
SQ Sequence 782 AA;

Query Match 49.3%; Score 231.5; DB 12; Length 782;
Best Local Similarity 48.3%; Pred. No. 5e-18;
Matches 42; Conservative 16; Mismatches 28; Indels 1; Caps 1;

QY 2 SGGVKVQYKNNDSAPGDNQIKPGQLVNTGSSSVDLSTVTVRYWFTDGGSSFLVYNCW 61
Db 631 sgdlavgyktgdtnaadngkphfnlvnkgaaavpilselrlyftad-gndqlqyncdw 689
QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 690 amvqcsnlngafvkmppgkanadtyle 716

RESULT 2
AAV13494
ID AAY13494 standard; Protein; 616 AA.
XX
AC AAY13494;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulase Cel E3/B5.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI; 1999-315403/27.
DR N-PSDB; AAX55660.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 42-43; 65pp; English.
XX

CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel B6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or

CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.
XX
SQ Sequence 616 AA;

Query Match 41.8%; Score 195.5; DB 20; Length 616;
Best Local Similarity 43.2%; Pred. No. 5.4e-14;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Caps 1;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGQLVNTGSSSVDLSTVTVRYWFTDGGSSFLVYNCW 60
Db 1 m9sgvkvlyknnetsastgsirpfwkivngsssvdlstrvkirwyvtdgdkpgsav-cd 59
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 60 waqigasvntfnfvklissgvgadyyile 87

RESULT 3
AAV13493
ID AAY13493 standard; Protein; 1751 AA.
XX
AC AAY13493;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulases comprising amino acid sequence.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI; 1999-315403/27.
DR N-PSDB; AAX55662.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 37-41; 65pp; English.
XX

CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel B6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic

XX	Sequence	1751 AA:
SQ		

Query Match 40.7%; Score 191.5; DB 20; Length 1426;
Best Local Similarity 43.5%; Pred. NO. 4.4e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1;

Qy 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSSVLDLSTVTVRWFTRDGGSSLVYNCDWAA 63
 :||| ||||::: |:| :|| | ||||| | :|||: || ||||
Dd 413 glkvlvknnetsastgsirpwfkilvnggsssvdlrvkirywtyvdgdkgpgasav-cdwaq 473

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
: | : : | : : |||:
Db 472 igasnvtfnfvklssqsvgadyytle 496

RESULT 5
AAR42122
ID AAR42122 standard; Protein; 499 AA.

AA AAR42122;

DT 27-APR-1994 (first entry)

DE NK-1 cellulase.

KW Cellulase; pH dependence; mutation.

PN JP05236969-A

PD 17-SEP-1993.

28-FEB-1992: 92JP-0075883.

XX
PR 28-FEB-1992: 92.TP-0075883XX
PA (REPP/) REPPU T

XX WPT: 1993-330585

DR N-PSDB; AAQ49820.
v v

PT Changing the pH-depe

XX
DP WPT: 1000-315103/27

DR WPI; 1999-315403/
DR N-PSDB: AAY55661

PS Disclosure; Fig 1; 9pp; Japanese.

XX

CC IS SHOWN IN sequence (AAQ43020). THE modelled NK-1 gene shows a
CC change in pH dependence.

Best Local Similarity 40.7%; Pred. No. 5.2e-13;

[illegible]

www.gtsvdykayasyvnsurqrprfintxtunngn

QY
63 AMGCGNIRASFGSVNPATPTADTYLQ

Db 410 qigcgnthkxfvltlhpkggadtyle 435

RESULT 6

AAR13227

ID AAR13227 standard; Protein; 700 AA.

XX

XX AAR13227;

XX AC

XX 14-OCT-1991 (first entry)

XX DT

XX Novel endoglucanase.

XX DE

XX Cellulase activity; detergent.

XX KW

XX OS

XX Bacillus spp. NCIMB 40250.

XX FH

XX Key Location/Qualifiers

XX Peptide 1..31

XX FT /label= signal sequence

XX FT 32..700

XX FT /label= mature endoglucanase

XX FT Cleavage-site 31..32

XX FT

XX PN WO9110732-A.

XX XX

XX 25-JUL-1991.

XX XX

XX 18-JAN-1991; 91WO-DK00013.

XX XX

XX 19-JAN-1990; 90DK-0000164.

XX XX

XX (NOVO) NOVO NORDISK A/S.

XX PA

XX Jorgensen PL, Schulein M, Hansen C;

XX PI

XX WPI; 1991-238020/32.

XX DR N-PSDB; AAR13001.

XX XX

XX Enzyme exhibiting cellulase activity from *Bacillus* sp. - is an

XX endo-glucanase, esp. useful for harshness redn. of cotton-contg.

XX PT fabrics.

XX PT

XX PS

XX Claim 1; Page 80; 96pp; English.

XX XX

XX The enzyme is produced by a strain of *Bacillus* spp. NCIMB 40250

XX and exhibits an endoglucanase activity of at least 10 (pref. at

XX least 25) carboxymethyl cellulose (CMC) endoase units per mg total

XX protein under alkaline conditions. It is especially useful as a

XX cellulolytic agent and has been found to be more stable during

XX washing (60 mins. at 40 deg.) in the presence of conventional

XX detergents than a commercial cellulase preparation. It may also

XX show increased storage stability in liq. detergents contg.

XX proteases. The sequence was deduced from the DNA (AAR13001).

XX See also AAR13228 and AAR13229.

XX SQ

Sequence 700 AA;

Query Match 36.6%; Score 172; DB 12; Length 700;

Best Local Similarity 44.9%; Pred. No. 3.3e-11;

Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKQYKND-SAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYNC 59

Db 549 vnsdlvvykgdgrnatndqkphfnkqkgtspvdlssltrryftkd-gsaamngwi 607

QY 60 DWAMCGNIRASFGSVNPATPTADTYLQ 88

Db 608 dwaklgsgniqisfghnga--dsdtyae 634

XX

RESULT 7

AAW18790

ID AAW18790 standard; protein; 551 AA.

XX

XX AAW18790;

XX AC

XX 18-NOV-1997 (first entry)

XX DT

XX Corrected *Bacillus lautus* (NCIMB 40250) endoglucanase Endo 3A.

XX DE

XX Endoglucanase; Endo 3A; formation; localised; variation;

XX KW colour density; surface; dye; fabric; family 5; cellulose;

XX KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;

XX KW blue jeans; back staining.

XX XX

XX *Bacillus lautus*.

XX OS

XX WO9709410-A1.

XX PN

XX 13-MAR-1997.

XX PD

XX 03-SEP-1996; 96WO-DK00364.

XX XX

XX 08-SEP-1995; 95DK-0000993.

XX PR

XX (NOVO) NOVO-NORDISK AS.

XX XX

XX Fich M, Onishi M, Schulein M, Toft AH;

XX PI

XX WPI; 1997-192888/17.

XX DR

XX Localised variation of colour density in the surface of a dyed

XX cellulosic fabric - uses cellulase compsn. able to hydrolyse

XX p-nitrophenyl -beta-1,4-cellobioside

XX PT

XX Disclosure; Pages 15-17; 23pp; English.

XX PS

XX The present sequence is the corrected version of the incorrect

XX *Bacillus lautus* (NCIMB 40250) endoglucanase Endo 3A described in

XX WO9110732. Endo 3 can be used in novel method of forming localised

XX colour density variation on the surface of a dyed cellulosic

XX fabric. The method comprises agitating the fabric in an aqueous

XX medium (pH 6.5 to 9.0) containing a family 5 cellulose,

XX e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta

XX -1,4-cellobioside, or a family 7 cellulase, and a mechanical

XX abrading agent or cellulose having abrading activity. Each

XX cellulase displays 30 % or more of its maximum activity at pH 7.

XX The process is useful to provide a stone washed look to blue jeans

XX without back staining.

XX Sequence 551 AA;

Query Match 35.2%; Score 165.5; DB 18; Length 551;

Best Local Similarity 37.9%; Pred. No. 1.4e-10;

Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVQYKND-SAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYNC 61

Db 402 tgnlvvykyvgdtsatdnqmkpsfnknngttptnlgsklryyftkd-gtadmsasfdw 460

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88

Db 461 aqigasnvsaaf--anftgsntdyve 485

XX

RESULT 8

AAR95080

ID AAR95080 standard; peptide; 167 AA.

XX

XX AAR95080;

XX AC

XX 30-OCT-1996 (first entry)

XX DT

XX

KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; bacillus sp. I633; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling
KW	subterranean formation fracture; cellulose binding domain.
XX	
OS	Synthetic.
OS	Bacillus sp.
OS	Clostridium thermocellum.
XX	
XX	WO9964619-A2.
PN	
XX	
PD	16-DEC-1999.
XX	
PF	10-JUN-1999; 99WO-DK00314.
XX	
PR	10-JUN-1998; 98US-0111256.
PR	20-OCT-1998; 98DK-0001340.
PR	20-OCT-1998; 98DK-0001341.
PR	28-OCT-1998; 98US-0105970.
PR	28-OCT-1998; 98US-0106054.
PR	23-DEC-1998; 98DK-0001725.
PR	05-MAR-1999; 99DK-0000306.
PR	05-MAR-1999; 99DK-0000307.
PR	05-MAR-1999; 99DK-0000308.
PR	05-MAR-1999; 99DK-0000309.
PR	09-MAR-1999; 99US-0123543.
PR	10-MAR-1999; 99US-0123623.
PR	10-MAR-1999; 99US-0123641.
PR	11-MAR-1999; 99US-0123642.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX	
DR	WPI; 2000-105891/09.
DR	N-PSDB; AA245336.
XX	
PT	New mannanases for treatment of textiles, plant material and coffee
PT	extract, and in cleaning compositions -
XX	
PS	Example 4; Page 211-212; 242pp; English.
XX	
CC	The present sequence represents a mannanase-linker-cellulose binding
CC	domain fusion protein. Mannanase (also known as mannan
CC	endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC	hydrolyses galactomannans. Specifically, mannanases hydrolyse
CC	1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans
CC	and galactoglucomannans. The mannanase protein, or preparations
CC	containing it, are used to improve properties of cellulosic or
CC	synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC	sizes or printing pastes). They are also used to degrade or modify
CC	plant materials (particularly recycled waste paper, paper making pulps,
CC	or material containing guar or locust bean gums (thickeners), or to
CC	reduce viscosity of mannan-containing foods or feeds). The mannanases
CC	are also used to process coffee extracts (to inhibit gel formation);
CC	in cleaning compositions (for machine washing of fabrics, as
CC	hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC	dental, contact lens or body-care compositions) where they remove
CC	mannan-containing soils and prevent binding of some soils to
CC	cellulotics; and in fabric softeners. They can also be used in oil
CC	well drilling to fracture subterranean formations.
XX	
XX	Sequence 476 AA;
SQ	

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSGSTLVYVNC 60

Db 314 vsgnlkvfyfynpsddttnsinpqfvtncgssaidlskltrlyyytvdgkqdtfw-cd 372

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 373 haaligsngsyngitsnvkgtfvkmsstnnadtyle 409

RESULT 10

AAAY28850
ID AAY28850 standard; Protein; 493 AA.

XX AC AAY28850;

XX DT 17-JAN-2000 (first entry)

DE DE Pectate lyase-linker-CBD fusion protein.

XX KW Pectate lyase-linker-CBD fusion protein; ATCC 14580; plasmid pMB914;
KW cellulose binding domain; CBD; linker; transformed; detergent;
KW cellulosic fibre; yarn; degradation; recycled waste paper; animal feed;
KW paper-making pulp; retting process; processing; wine; juice.

XX OS Chimeric - Bacillus licheniformis.

XX OS Chimeric - Clostridium thermocellum.

XX PN WO9927083-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00514.

XX PR 24-NOV-1997; 97DK-0001344.

XX PR 06-MAY-1998; 98US-0073684.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;

XX DR WPI; 1999-610578/52.

XX DR N-PSDB; AAX90978.

XX PT New isolated pectate lyase enzymes -

XX PS Example 4; Page 85-86; 93pp; English.

XX CC The present sequence is a Pectate lyase-linker-CBD fusion protein.

CC CC Plasmid pMB914 was constructed using pectate lyase gene from

CC Bacillus licheniformis, ATCC 14580 and cellulose binding domain (CBD)

CC gene from Clostridium thermocellum YS fused with a linker molecule.

CC CC Bacillus subtilis was transformed with plasmid pMB914 for expression of

CC the fusion protein. Pectate lyase can be used in detergent compositions,

CC for cleaning hard surfaces, for machine treatment of fabrics, for

CC improving the properties of cellulosic fibres, yarn, woven or non-woven

CC fabric, for the degradation of plant material e.g. recycled waste paper,

CC mechanical paper-making pulps or fibres subjected to retting process, for

CC preparing animal feed and for processing wine or juice.

XX SQ Sequence 493 AA;

Query Match 35.1%; Score 165; DB 20; Length 493;

Best Local Similarity 36.1%; Pred. No. 1.4e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSGSTLVYVNC 60

Db 331 vsgnlkvfyfynpsddttnsinpqfvtncgssaidlskltrlyyytvdgkqdtfw-cd 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 390 haaligsngsyngitsnvkgtfvkmsstnnadtyle 426

RESULT 11

AAAY43218
ID AAY43218 standard; Protein; 493 AA.

XX AC AAY43218;

XX DT 13-JAN-2000 (first entry)

XX DE Pectate lyase CBD fusion protein sequence.

XX KW Pectate lyase; polysaccharide lyase; enzyme; pectin degradation;
KW polygalacturonide; detergent composition; hard surface treatment;
KW cellulosic fibre; plant material degradation; recycled waste paper;
KW mechanical paper-making pulp; wine processing; cellulose binding domain;
KW CipB.

XX OS Clostridium thermocellum.

OS OS Bacillus sp.

OS OS Synthetic.

XX PN WO9927084-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00515.

XX PR 24-NOV-1997; 97DK-0001343.

XX PR 06-MAY-1998; 98US-0073684.

XX PR 02-NOV-1998; 98US-0184217.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Schuelein M, Lange NEK, Bjornvad ME, Moller S;

XX PI Glad SOS, Kauppinen MS, Schnorr K, Kongsbaek L;

XX DR WPI; 1999-610579/52.

XX DR N-PSDB; AAX31562.

XX PT New isolated pectate lyase enzymes -

XX PS Claim 26; Page 106-108; 113pp; English.

XX CC This sequence is a fusion protein comprising a Bacillus species
CC pectate lyase of the invention, fused via a linker to the Clostridium
CC thermocellum CipB protein cellulose binding domain (CBD). The pectate
CC lyase enzymes are obtained from Bacillus agaradhaerens (BA), Bacillus
CC licheniformis (BL), Bacillus halodurans (BH) and other Bacillus species.
CC The pectate lyase enzymes can be used for degrading pectin, pectate and
CC polygalacturonides. They can be used in detergent compositions, for
CC cleaning hard surfaces, for machine treatment of fabrics, for improving
CC the properties of cellulosic fibres, yarn, woven or non-woven fabric, for
CC the degradation of plant material (e.g. recycled waste paper, mechanical
CC paper-making pulps or fibres subjected to a retting process, for
CC preparing animal feed or for processing wine or juice. DNA encoding the
CC enzymes can also be used for the production of transgenic plants.

XX SQ Sequence 493 AA;

Query Match 35.1%; Score 165; DB 20; Length 493;

Best Local Similarity 36.1%; Pred. No. 1.4e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSGSTLVYVNC 60

Db 331 vsgnlkvfyfynpsddttnsinpqfvtncgssaidlskltrlyyytvdgkqdtfw-cd 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Search completed: July 2, 2002, 09:15:51
Job time: 239 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:54 ; Search time 39.87 Seconds
(without alignments)
54.524 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	616	4	US-09-136-574A-47
2	194.5	41.4	1751	4	US-09-136-574A-44
3	191.5	40.7	1426	4	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PCT-US95-13813-9
7	165	35.1	493	4	US-09-198-956-10
8	165	35.1	493	4	US-09-198-955A-12
9	120.5	25.6	531	2	US-07-862-588B-7
10	114	24.3	162	1	US-08-048-164A-2
11	114	24.3	162	1	US-08-460-462-2
12	114	24.3	162	1	US-08-460-457-2
13	114	24.3	162	1	US-08-460-458-2
14	114	24.3	162	2	US-08-460-455-2
15	114	24.3	162	2	US-08-330-394A-2
16	114	24.3	163	3	US-09-006-636-7
17	114	24.3	163	4	US-09-006-632-7
18	113	24.0	382	4	US-09-277-716-22
19	112	23.8	154	2	US-08-330-394A-29
20	112	23.8	156	2	US-08-330-394A-22
21	64	13.6	428	3	US-09-118-319-5
22	64	13.6	464	1	US-08-353-400-36
23	63.5	13.5	1785	4	US-09-341-587-3
24	63	13.4	1581	4	US-09-110-517-2
25	62.5	13.3	288	4	US-09-423-439-38
26	62.5	13.3	445	1	US-08-353-400-33
27	62.5	13.3	673	4	US-09-423-439-32

28	62.5	13.3	802	4	US-09-081-345-18	Sequence 18, Appl
29	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appl
30	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
31	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
32	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
33	61.5	13.1	307	4	US-08-370-476-68	Sequence 68, Appl
34	61.5	13.1	316	1	US-08-260-202A-18	Sequence 18, Appl
35	61.5	13.1	316	1	US-08-017-114-18	Sequence 18, Appl
36	61.5	13.1	316	3	US-08-505-307-18	Sequence 18, Appl
37	61.5	13.1	316	5	PCT-US94-02034-18	Sequence 18, Appl
38	61	13.0	301	2	US-08-656-906-25	Sequence 25, Appl
39	61	13.0	301	4	US-09-217-847-25	Sequence 25, Appl
40	61	13.0	617	1	US-08-361-920-29	Sequence 29, Appl
41	61	13.0	617	1	US-08-479-939-29	Sequence 29, Appl
42	61	13.0	617	1	US-08-483-432-29	Sequence 29, Appl
43	60.5	12.9	240	1	US-08-488-113B-147	Sequence 147, App
44	60.5	12.9	240	1	US-08-477-484B-147	Sequence 147, App
45	60.5	12.9	240	2	US-08-646-360-147	Sequence 147, App

ALIGNMENTS

RESULT 1
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland

Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA

COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:


```

; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 35.2%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 2.2e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2

Qy 2 SGGVKVQYKKNDSAPGNDQIKPGQLVNTGSSSDLSSTVTVRYWFTRDGSGSTLYNCW 61
Db 402 TGNLVQYKVGDTSATDNQMKFSFNKNGTTPVNLISGLKLYRYTKD-GTADMSASFDW 460
Qy 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
Db 461 AQIGASNVSAF--ANFTGSNTDTYVE 485

RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brodsky and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington

```


STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13813
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.1%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 5.7e-11;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSDAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 5 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQKQDTFW-CD 63
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 64 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 100

RESULT 7
US-09-198-956-10
Sequence 10, Application US/09198956
Patent No. 6165769
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schulein, Martin
APPLICANT: Lange, Niels Erik K.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Schnorr, Kirk
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
FILE OF INVENTION: Licheniformis
FILE REFERENCE: 5377.200-US
CURRENT APPLICATION NUMBER: US/09/198,956
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: 1344/97
EARLIER FILING DATE: 1997-11-24
EARLIER APPLICATION NUMBER: 60/067,240
EARLIER FILING DATE: 1997-12-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 493
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSDAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 331 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 390 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 8
US-09-198-955A-12
Sequence 12, Application US/09198955A
Patent No. 6187580
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schulein, Martin
APPLICANT: Lange, Niels E.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Moller, Soren
APPLICANT: Glad, Sanne O. S.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars
TITLE OF INVENTION: NO. 6187580el Pectate Lyases
FILE REFERENCE: 5378.200-US
CURRENT APPLICATION NUMBER: US/09/198,955A
CURRENT FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067,249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067,240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073,684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184,217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSDAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 331 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 390 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 9
US-07-862-588B-7
Sequence 7, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnaa
APPLICANT: Schulein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7


```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
;
; PRIOR APPLICATION DATA: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-7

Query Match 25.6%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. NO. 1.9e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCW 61
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 432 TGNLVQYKVGDSATDNQKPFNIKNGTTPVNLGLKXXXXKD-GPADMSCSIDW 490
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||

QY 62 AMGCGNIRASFGSVNPATPTADTY 86
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 491 AQIGRTNVLLAF--ANFTGNTDITY 513
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||

RESULT 10
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiel, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
```

```
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-048-164A-2

Query Match 24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 2.3e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCW 61
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 3 TSSMVEFYNSNKSQNTSITPIKTIKNTSDSLNLDNVRVYVYTSDGQGQTFW-CDH 61
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||

QY 62 AMGCGN-----IRASF--GSVNPATPTADTYLQ 88
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97
: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||

RESULT 11
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiel, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	191	40.6	1711	2	T31337	1,4-beta-glucanase	
2	190.5	40.5	1779	2	T31085	xylanase - Caldice	
3	190	40.4	508	2	G9593	cellulase (EC 3.2.	
4	190	40.4	508	2	A26874	cellulase (EC 3.2.	
5	188	40.0	499	2	JN0111	cellulase (EC 3.2.	
6	186.5	39.7	1331	2	A48954	mannan endo-1,4-be	
7	186.5	39.7	1742	2	T17120	cellulase (EC 3.2.	
8	186	39.6	499	2	A27198	cellulase (EC 3.2.	
9	182.5	38.8	915	2	A43802	cellulase (EC 3.2.	
10	182.5	38.8	1039	2	S02711	cellulase (EC 3.2.	
11	180.5	38.4	145	2	A41897	cellulase homolog	
12	180	38.3	486	2	I40548	bifunctional cellu	
13	172	36.6	700	2	B41897	cellulase (EC 3.2.	
14	165.5	35.2	879	2	A47704	endoglucanase I (E	
15	165	35.1	1854	2	S36859	clpA protein - Clo	
16	149.5	31.8	505	2	S39962	cellulase - Er	
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2.	
18	141.5	30.1	986	2	S12021	thermoactive cellu	
19	130.5	27.8	586	2	PC6006	scaffolding protei	
20	130.5	27.8	1483	2	C37012	probably celluloso	
21	126.5	26.9	1162	2	T30433	scaffolding protei	
22	117	24.9	1230	2	S47466	cellulose 1,4-beta	
23	114	24.3	1848	2	A44140	cellulose-binding	
24	75.5	16.1	618	2	T08685	hypothetical prote	
25	73.5	15.6	547	2	T25478	hypothetical prote	
26	73.5	15.6	1428	2	AG2224	hypothetical prote	
27	70	14.9	5188	2	B85347	probable RTX famil	
28	70	14.9	5291	2	T69096	hypothetical prote	
29	68.5	14.6	574	2	A69196	cell surface glyco	

A; Note: xynC

Query Match	40.5%	Score 190.5;	DB 2;	Length 1779;
Best Local Similarity	43.5%;	Pred. NO. 1.4e-11;		
Matches 37;	Conservative 14;	Mismatches 33;	Indels 1;	Gaps 1;

A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <WOL>
A:CROSS-references: EMBL:D29076; NID:g509266; PIDN:CRA82317.1; PID:g509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
Korean J. Microbiol. 24, 236-242, 1986
A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
A:Reference number: I39803
A:Accession: I39803
A>Status: preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: DNA
A:Residues: 10-13,'V','I5','N','17-18','V','20-21','F','23','A','25-26','AI','29-31','P','33','PO',
A:CROSS-references: GB:M38634; NID:g142657; PIDN:AAA22300.1; PID:g142658
A:Experimental source: strain ATCC 6633
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
G:Genetics:
A:Gene: bgic
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
E:I-38/Domain: signal sequence #status predicted <SIG>
F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.4%; Score 190; DB 2; Length 508;
Best Local Similarity 40.7%; Pred. No. 4.3e-12;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

Qy 4 GVKYVKNNDSAPGDNIKPGQLVNTGGSSVDLSVTYRYWF--TRDGGSGSTLYNCDA 62
| : |||| | : |||| | : ||::|| || :: | : : ||::|
Db 362 GISVOYRAGDGSMNSNQIRPOLQIKNGNTVDLKDTARYWKAKNKGN---FDGDYA 418

Qy 63 AMGCCNRASFGSVNPATPTADTYLQ 88
:||||| | : |||||
Db 419 QIGCGNVTHKEVTLHKPKQGADTYLE 444

RESULT 4
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1-4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:CROSS-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A>Note: the authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; po
F:I-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.4%; Score 190; DB 2; Length 508;
Best Local Similarity 43.0%; Pred. No. 4.3e-12;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

Qy 4 GVKYVKNNDSAPGDNIKPGQLVNTGGSSVDLSVTYRYWF--TRDGGSGSTLYNCDA 62

C; Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.7%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVOYKKNDSAPGDNQIKPGQLVNTSGSSVDLSVTVRVWFTRDGGSSTLVNCDW 61
|| || || : : | : || : || : || : || : || : || : || : || : ||
Db 364 SGQIKVLVYANKETNSTTIRPWLKVNVSGSSSIDLSRVTIRYWTVTDGERAQAIS-DW 422

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
| : | | : | : : : | | ||:
Db 423 AQIGASNVTFFVKLVSSVSAGADYYLE 449

RESULT 7

Tl7120
cellulase (EC 3.2.1.1) precursor, thermoactive - Caldocellum saccharolyticum
C; Species: Caldocellum saccharolyticum
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C; Accession: Tl7120; AA3745
R; Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A; Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophile
A; Reference number: Z18698; MUID:95336703
A; Accession: Tl7120
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-1742 <TEO>
A; Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R; Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A; Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a multidomain cellulase from the extreme thermophile
A; Reference number: A43745; MUID:91247819
A; Accession: A43745
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1516-1544, 'A' 1546-1742 <LUE>
A; Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C; Genetics:
A; Gene: cea
C; Keywords: glycosidase; hydrolase

Query Match 39.7%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVOYKKNDSAPGDNQIKPGQLVNTSGSSVDLSVTVRVWFTRDGGSSTLVNCDW 61
|| || || : : | : || : || : || : || : || : || : || : || : ||
Db 704 SGQIKVLVYANKETNSTTIRPWLKVNVSGSSSIDLSRVTIRYWTVTDGERAQAIS-DW 762

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
| : | | : | : : : | | ||:
Db 763 AQIGASNVTFFVKLVSSVSAGADYYLE 789

RESULT 8

AZ7198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IF03034)
N; Alternate names: endo-1,4-beta-glucanase
C; Species: Bacillus subtilis
C; Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C; Accession: AZ7198
R; Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A; Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A; Reference number: A27198; MUID:87190397
A; Accession: A27198
A; Molecule type: DNA
A; Residues: 1-499 <NAK>
A; Cross-references: GB:M28332; NID:q142670; PIDN:AAA22307.1; PID:q142671

A:Experimental source: strain IF03034
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
 A:Pathway: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 39.6%; Score 186; DB 2; Length 499;
 Best Local Similarity 40.7%; Pred. No. 1.1e-11;
 Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

QY 4 GGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWF-TRDGGSSTLVYNCDDWA 62
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 353 GGVKVOYKAGDGNVNSQIRPOLHKNNGNATVDLKDVTARYWYNAKNGQN---FDCDYA 409

QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 410 QIGCGNLTHKFTVLHKKPGQADTYLE 435

RESULT 9
 A43802
 cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum s
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Caldocellum saccharolyticum
 C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
 C:Accession: A43802
 R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 56: 3117-3124, 1990
 A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile
 A:Reference number: A43802; MUID:91136262
 A:Accession: A43802
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: DNA
 A:Residues: 1-915 <SAU>
 A:Cross-references: EMBL:X13602
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A:Pathway: cellulose degradation
 C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 915;
 Best Local Similarity 41.9%; Pred. No. 4.8e-11;
 Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVYNCDDWA 62
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 349 GOIKVLYANKETNSTTIRPWLKVVNSGSSIDLSTLRVIRYWTVDGERAQSAYS-DWA 407

QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 408 QIGASNVTFFKVLSSVSGADTYLE 433

RESULT 10
 S02711
 cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
 N:Alternate names: endo-1,4-beta-glucanase
 N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
 C:Species: Caldocellum saccharolyticum
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
 C:Accession: S02711
 R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
 Nucleic Acids Res. 17, 439, 1989
 A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo
 A:Reference number: S02711; MUID:89098398
 A:Accession: S02711
 A:Molecule type: DNA
 A:Residues: 1-1039 <SAU>

A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
 C:Genetics:
 A:Gene: celB
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-1039/Product: cellulase #status predicted <MAT>
 F:172-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 1039;
 Best Local Similarity 41.9%; Pred. No. 5.6e-11;
 Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCDDWA 62
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 420 GOIKVLYANKETNSTTIRPWLKVVNSGSSIDLSTLRVIRYWTVDGERAQSAYS-DWA 478

QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 479 QIGASNVTFFKVLSSVSGADTYLE 504

RESULT 11
 A41897
 cellulase homolog - Bacillus lautus (fragment)
 C:Species: Bacillus lautus
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
 C:Accession: A41897; S27498
 R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
 J. Bacteriol. 174, 3522-3531, 1992
 A:Title: celA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-
 A:Reference number: A41897; MUID:92276330
 A:Accession: A41897
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <HAN>
 A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662
 A:Experimental source: PL236
 A:Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 38.4%; Score 180.5; DB 2; Length 145;
 Best Local Similarity 43.9%; Pred. No. 1.1e-11;
 Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

QY 7 VOYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCDDWAMGC 66
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 1 LOYRAADTNADNQIKPSFNKNGTSADVDSLTKIRYFTKDGSAAVNGW-IDNAQLGG 59

QY 67 GNIRASFGSVNPATPTADTYLQ 88
 I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: |||| I: ||||
 Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 12
 I40548
 bifunctional cellulase precursor - Bacillus sp.
 C:Species: Bacillus sp.
 C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
 C:Accession: I40548
 R:Han, S.-J.; Yoo, Y.J.; Kang, H.S.
 J. Biol. Chem. 270, 26012-26019, 1995
 A:Title: Characterization of a bifunctional cellulase and its structural gene: the ce
 A:Reference number: I40548; MUID:96029707
 A:Accession: I40548
 A:Status: preliminary; translated from GB/EMBL/DBDJB
 A:Molecule type: DNA
 A:Residues: 1-486 <RES>
 A:Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576

Qy 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 424 HAAIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: July 2, 2002, 09:14:14
Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:18:03 ; Search time 28.05 Seconds
(without alignments)
122.853 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	190	40.4	499	1 GUN1_BACSU	P07983 bacillus su
2	190	40.4	499	1 GUN2_BACSU	P10475 bacillus su
3	188	40.0	499	1 GUN3_BACSU	P23549 bacillus su
4	186.5	39.7	1331	1 MANB_CALSA	P22533 caldocellum
5	186.5	39.7	1742	1 GUNA_CALSA	P22534 caldocellum
6	182.5	38.8	1039	1 GUNB_CALSA	P10474 c endogluca
7	180.5	38.4	145	1 YCEA_PAECLA	P29718 paenibacill
8	172	36.6	700	1 GUNA_PAECLA	P29719 paenibacill
9	165.5	35.2	879	1 GUN1_CLOTM	Q02934 clostridium
10	165	35.1	772	1 CIPB_CLOTM	Q01866 clostridium
11	165	35.1	1853	1 CIPA_CLOTM	Q06851 clostridium
12	149.5	31.8	444	1 GUNN_ERWCA	Q59394 erwinia car
13	149.5	31.8	505	1 GUNV_ERWCA	Q47096 erwinia car
14	147.5	31.4	504	1 GUNW_ERWCA	Q59395 erwinia car
15	147.5	31.4	914	1 GUX2_CLOSR	P50900 clostridium
16	141.5	30.1	986	1 GUNZ_CLOSR	P23659 clostridium
17	114	24.3	1848	1 CBPA_CLOCL	P38058 clostridium
18	67.5	14.4	118	1 HV39_MOUSE	P01809 mus musculu
19	67.5	14.4	215	1 FLA1_METVA	P95316 methanococc
20	65	13.8	523	1 PUR9_MYCTU	P71553 m bifunctio
21	64	13.6	1571	1 ATC5_YEAST	P32660 saccharomyc
22	63.5	13.5	247	1 FLA1_THEVO	P57719 thermoplasm
23	63.5	13.5	465	1 TY3H_SCHMA	O17446 schistosoma
24	63	13.4	953	1 YA42_HUMAN	Q19699 homo sapien
25	63	13.4	1581	1 PPRB_HUMAN	O15648 h peroxisom
26	62.5	13.3	269	1 EXOK_RHIME	P13693 rhizobium m
27	62.5	13.3	802	1 PTNB_MOUSE	P29352 mus musculu
28	62	13.2	1210	1 ICEN_PSEFL	P09815 pseudomonas
29	61.5	13.1	252	1 HIS6_RHOSH	P50937 rhodobacter
30	61.5	13.1	316	1 MUCB_PSEAE	P38108 pseudomonas
31	61.5	13.1	328	1 HAIQ_MOUSE	P14428 mus musculu
32	61.5	13.1	368	1 HAIW_MOUSE	P03991 mus musculu
33	61.5	13.1	678	1 YIHO_ECOLI	P32138 escherichia

ALIGNMENTS

RESULT 1

ID	GUN1_BACSU	STANDARD;	PRT;	499 AA.
AC	P07983;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)			
DE	(Cellulase).			
GN	BGLC OR GLD.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DLG;			
RX	MEDLINE=87194581; PubMed=3106328;			
RA	Robson L.M., Chambliss G.H.;			
RT	"Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";			
RL	J. Bacteriol. 169:2017-2025(1987)			
CC	- - CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.			
CC	- - SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M16185; AAA22496.1; ALT_INIT.			
DR	PIR; A26874; A26874.			
DR	HSSP; O85465; 1A3H.			
DR	InterPro: IPR001956; CBD_3.			
DR	InterPro: IPR001547; Glyco_hydro_F5.			
DR	Pfam; PF00942; CBD_3; 1.			
DR	Pfam; PF00150; cellulase; 1.			
DR	ProDom; PD001947; CBD_3; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL 1 29			
FT	CHAIN 30 499			
FT	ENDOGLUCANASE.			
FT	ACT_SITE 169 169			
FT	PROTON DONOR (BY SIMILARITY).			
FT	ACT_SITE 257 257			
FT	NUCLEOPHILE (BY SIMILARITY).			
FT	DOMAIN 350 499			
FT	CELLULOSE-BINDING (BY SIMILARITY).			
SQ	SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;			

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 43.0%; Pred. No. 4.4e-14;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;


```
QY 4 GVKVQYKNDSDAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 GSVQYKAGDGRVNSQIRPQLHKNGNATVDLKDVTARYWYVNVKNGQN---FDCDYA 409

QY 63 AMCGCNIRASFSGSVNPATPTADTYLQ 88
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 410 QMCGGNLTHKEVTLHKPKQGADTYLE 435

RESULT 2
GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
GN BGLC OR GLD OR EGLS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAP115;
RX MEDLINE=87066783; PubMed=3024130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moranelli F., Selligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rose M., Entian K.;
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Fiengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; CAA82317.1; -
DR EMBL; X04689; CAA28392.1; -
DR EMBL; X67044; CAA47429.1; -
DR EMBL; Z73234; CAA97610.1; ALT_INIT.
DR EMBL; Z99113; CABI3696.1; ALT_INIT.
DR PIR; A26114; A26114.
DR HSSP; O85465; 1A3H.
```

```
DR Subtilist; BG10437; bglc.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29 ENDOGLUCANASE.
FT CHAIN 30 499 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 169 169 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAE2 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 4.4e-14;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSDAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 GSVQYKAGDGRVNSQIRPQLHKNGNATVDLKDVTARYWYVNVKNGQN---FDCDYA 409

QY 63 AMCGCNIRASFSGSVNPATPTADTYLQ 88
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 410 QMCGGNLTHKEVTLHKPKQGADTYLE 435

RESULT 3
GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCcase) (Cellulase).
GN BGLC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BSE616;
RX MEDLINE=91299280; PubMed=1368694;
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616.";
RL Agric. Biol. Chem. 55:441-448(1991).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01057; BAA00859.1; -
DR PIR; JN0111; JN0111.
DR HSSP; O85465; 1A3H.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00150; cellulase; 1.
```



```
DR ProDom: PD001947; CBD_3; 1.
KW PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55169 MW; 2E821E3D8BACA04 CRC64;

Query Match 40.0%; Score 188; DB 1; Length 499;
Best Local Similarity 40.0%; Pred. No. 7 4e-14;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 4 GVKVQYKNDSPGDNOIKPGLQVNTGSSVDLSTVTVRYWFTDGGSSLTLYNCWAA 63
DB 353 GISVQYRAGDGMNSQIRPOLQIKNGNTVTDKDVTRYWYNAKNGQNV--DCDYAQ 410

QY 64 MCGCNIRASFGSVNPTPTADTYLQ 88
DB 411 LGGCNVYKFTVTLHKPKQGADTYLE 435

RESULT 4
MANB_CALSA STANDARD; PRT; 1331 AA.
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
DE mannanase) (Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase))].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93119139; PubMed=1476429;
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme.";
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luethi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01257; AAA71887.1; -.
DR EMBL; M36063; AAA72861.1; -.
DR PIR; B43745; B43745.
DR PIR; A48954; A48954.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; Glyco_hydro_F5.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Hydrolase; Glycosidase; Cellulose degradation; Signal;
KW Multifunctional enzyme.
FT SIGNAL 1 41
FT CHAIN 42 1331
FT DOMAIN 42 325
FT DOMAIN 326 361
FT DOMAIN 362 518
FT DOMAIN 519 564
FT DOMAIN 565 720
FT DOMAIN 721 780
FT DOMAIN 781 1331
FT ACT_SITE 162 162
FT ACT_SITE 257 257
FT CONFLICT 338 338
FT CONFLICT 340 346
SQ SEQUENCE 1331 AA; 146892 MW; PFBCA51BB8D8F0E0 CRC64;

Query Match 39.7%; Score 186.5; DB 1; Length 1331;
Best Local Similarity 42.5%; Pred. No. 3.5e-13;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSPGDNOIKPGLQVNTGSSVDLSTVTVRYWFTDGGSSLTLYNCW 61
DB 364 SGOIKVLYANKSTNTTIRPWLKVVNSGSSSIDLSRVTRYWYTVVDGERAQSATS-DW 422

QY 62 AAMCGCNIRASFGSVNPTPTADTYLQ 88
DB 423 AQIGASNVTFKFKVLSVSGADYILE 449

RESULT 5
GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
```


QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
Db 479 QIGASNVTFKFKLSVSSVSGADYILE 504

RESULT 7
YCEA_PAELA STANDARD; PRT; 145 AA.
AC P29718;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in CELA 5' region (Fragment).
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Joergensen P.L., Diderichsen B.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase".
RL J. Bacteriol. 174:3522-3531(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76588; AAA22303.1; -
CC PIR: A41897; A41897.
CC PIR: S27498; S27498.
CC HSSP: Q06851; INBC.
CC InterPro: IPR001956; CBD_3.
CC Pfam: PF00942; CBD_3; 1.
CC ProDom: PD001947; CBD_3; 1.
CC Hypothetical protein.
KW NON_TER 1
FT SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;
SQ

Query Match 38.4%; Score 180.5; DB 1; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.2e-13;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

QY 7 VOYKNDSAPGDNQIRKGLQVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCDAWAMGC 66
Db 1 LQYRAADTNAADNQIRKPSFNKNGTSVAVDLSTLKIRYFTKDSAAVNGW-IDWAQLGG 59
QY 67 GNTIRASFGSVNPATPTADTYLQ 88
Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 8
GUNA_PAELA STANDARD; PRT; 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.

OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase".
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76588; AAA22303.1; -
CC PIR: B41897; B41897.
CC PIR: S27499; S27499.
CC HSSP: Q06851; INBC.
CC InterPro: IPR001956; CBD_3.
CC Pfam: PF00942; CBD_3; 1.
CC ProDom: PD001947; CBD_3; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 700 ENDOGLUCANASE A.
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 213 213 BY SIMILARITY.
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EE0F CRC64;
SQ

Query Match 36.6%; Score 172; DB 1; Length 700;
Best Local Similarity 44.9%; Pred. No. 7.4e-12;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKNNDSAPGDNQIRKGLQVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNC 59
Db 549 VNSDLVVQYKDGDRNNATDNQIKPFHNIQNKTSVDLSTLRYFTKD-SSAAMNGWI 607
QY 60 DWAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 608 DWAKLGSNIQISFGNHGA--DSDIYAE 634

RESULT 9
GUNI_CLOTM STANDARD; PRT; 879 AA.
AC Q02934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)
DE (Cellulase I).
GN CELI.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
RC STRAIN=NCIB 10682;
RX MEDLINE=93171873; PubMed=8436949;
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
RT Gilbert H.J.;
*Gene sequence and properties of CellI, a family E endoglucanase from


```
DR InterPro: IPR000556; Glyco_hydro_48.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR PRINTS: PR00844; GLHYDLASE48.
DR PRODOM: PD001947; CBD_3; 1.
DR PRODOM: PD011903; Glyco_hydro_48; 1.
DR PRODOM: PD011903; Glycosidase; Signal.
DR KW Cellulose degradation; Hydrolase; Signal.
FT SIGNAL 1 33 POTENTIAL
FT CHAIN 34 914 EXOGLUCANASE II..
SQ SEQUENCE 914 AA; 103020 MW; D0DB6017D6DF82C CRC64;

Query Match 31.4%; Score 147.5; DB 1; Length 914;
Best Local Similarity 35.2%; Pred. No. 6.2e-09;
Matches 31; Conservative 21; Mismatches 35; Indels 1; Gaps

Qy 1 VSGGVKVOYKNDSAPCDNOIKCLQLVNTGSSVDLSTVTVRWFTRDCGSSLTVYNCD 60
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||::||
Db 763 VEGVLIIQSFNANTFOETNSIMPRFRINSGNITSIPLEVKLRYYITVGDGKPNFW-CD 821
||||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Qy 61 WAAMCGGNIRASFGSVNPATPTADTYIQ 88
||||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Db 822 WASIGSNVVTGTFFVKMDGATTGADYYLE 849
||||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
```

Search completed: July 2, 2002, 09:18:04
Job time: 372 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:33 ; Search time 91.42 Seconds
(without alignments)
168.416 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviro.*
- 16: sp_bacteriap.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	225.5	48.0	170	2	Q9RFX6	Q9rfx6 caldibacill
2	216.5	46.1	930	2	Q9REX5	Q9rfx5 caldibacill
3	214.5	45.6	921	2	Q9L8L8	Q9l8l8 caldibacill
4	204.5	43.5	987	2	Q9Z4I1	Q9z4i1 bacillus sp
5	194.5	41.4	1751	2	Q9A0G4	Q9a0g4 caldicellul
6	192.5	41.0	1000	2	Q9A0G4	Q9a0g4 caldicellul
7	192.5	41.0	1770	2	Q9X3P5	Q9x3p5 caldicellul
8	191.5	40.7	261	2	Q9A0G7	Q9a0g7 caldicellul
9	191.5	40.7	1426	2	Q9X3P6	Q9x3p6 caldicellul
10	191	40.6	1711	2	P96311	P96311 anaerocellu
11	190.5	40.5	996	2	Q9AQH0	Q9aqh0 caldicellul
12	190.5	40.5	1779	2	Q52374	Q52374 caldicellul
13	190	40.4	499	2	Q93TJ6	Q93tj6 bacillus su
14	190	40.4	508	2	Q93LD0	Q93ld0 bacillus su
15	187	39.8	499	2	O52731	O52731 bacillus sp
16	186	39.6	499	2	Q45532	Q45532 bacillus su

17	186	39.6	501	2	O83012	O83012 bacillus sp
18	180	38.3	486	2	Q45430	Q45430 bacillus sp
19	172.5	36.7	1915	2	Q9REL0	Q9rpl0 acetivibrio
20	168	35.7	473	2	Q9RK75	Q9rk75 streptomyce
21	165.5	35.2	887	2	Q9L3J8	Q9l3j8 clostridium
22	150.5	32.0	2316	2	Q9FDJ9	Q9fdj9 bacteroides
23	130.5	27.8	1493	16	Q977Y4	Q977y4 clostridium
24	130.5	27.8	1546	2	Q45996	Q45996 clostridium
25	126.5	26.9	1162	2	O82830	O82830 clostridium
26	117	24.9	1230	2	Q59325	Q59325 clostridium
27	116.5	24.8	321	2	Q46392	Q46392 clostridium
28	87	18.5	440	2	Q93LI9	Q93li9 bacillus sp
29	78	16.6	135	13	Q90542	Q90542 ginglymosto
30	75.5	16.1	618	4	Q9Y3Z2	Q9y3z2 homo sapien
31	75.5	16.1	660	2	Q9L3J2	Q9l3j2 clostridium
32	75.5	16.1	986	4	Q94858	Q94858 homo sapien
33	75.5	16.1	1346	4	Q9UIZ3	Q9uiz3 homo sapien
34	73.5	15.6	547	5	P91006	P91006 caenorhabdi
35	73	15.5	511	5	Q9GYG5	Q9gyg5 caenorhabdi
36	72	15.3	499	12	Q993M3	Q993m3 autonomous
37	72	15.3	582	12	P90329	P90329 kilham rat
38	72	15.3	639	12	Q993M4	Q993m4 autonomous
39	72	15.3	722	12	P88901	P88901 kilham rat
40	71.5	15.2	263	17	Q9HID3	Q9hid3 thermoplasm
41	71	15.1	389	3	Q9Y775	Q9y775 candida tro
42	70	14.9	137	13	Q9YHP9	Q9yhp9 ginglymosto
43	69.5	14.8	1430	16	Q98HK1	Q98hki rhizobium l
44	69	14.7	2873	12	O93073	O93073 hepatitis g
45	68.5	14.6	574	17	O26815	O26815 methanother

ALIGNMENTS

RESULT 1

Q9RFX6 PRELIMINARY; PRT; 170 AA.
AC Q9RFX6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN (FRAGMENT).
OS Caldicoccus cellulosivorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicoccus cellulosivorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; ACP22273.1; -.
DR HSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
FT Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F4E3A0B4 CRC64;

Query Match 48.0%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred No. 2.6e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;
Qy 3 GGVQYKNNDSAPGDNQIKPGLQVNTGSSSVLSTVTYVWFTRDGGSTLVNCDWA 62
I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I :
Db 20 GSLVQVYRAADTNAGDNLKPHFRIVNRTGSSVPLSELTIRYWTVD-GDKPQVFNCDWA 78
Qy 63 AMCCGNTRASFGSVNPATPTADTYLQ 88

DE	BETA-GLUCANASE.
OS	thermophilic anaerobe NA10.
OC	Bacteria.
ON	NCBI_TaxID=67756;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NA10;
RA	Miyake K., Machida Y., Hattori K., Iijima S.;
RT	"Characterization of a multi-domain cellulase from an extremely
RL	thermophilic anaerobe strain NA10.";
DR	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AS008029; BAA22939.1; -
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001589; Actinin_act_bind.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR001000; Glyco_hydro_10.
DR	InterPro; IPR001547; Glyco_hydro_F5.
DR	Pfam; PF00942; CBD_3; 1.
DR	Pfam; PF00150; cellulase; 1.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS; PR00134; GLHYDRLASE10.
DR	ProDom; PD001947; CBD_3; 1.
DR	PROSITE; PS00019; ACTININ.1; UNKNOWN1.
DR	PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN1.
SQ	SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;
Query Match	41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity	43.7%; Pred. No. 9.9e-13;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps	
QY	2 SGGVKYQYNKNSAPGDNIKPGQLQLVNTGSSVSLSTVRYWFTRDGSGSLTVYNCDW 61 : : : : : : : : : :
Db	372 SGQIKVLYANKETNSTNTIRPLWKVVNTGSSSIDLSRVITRWYTVDGKAQSAYS-DW 430
QY	62 AAMCGGNIRASFGSVNPATPTADTYLQ 88 : : : : : : :
Db	431 AQIGASNVTFFKVKLSSVSAGADYYLE 457
RESULT 7	
ID Q9X3P5	PRELIMINARY; PRT; 1770 AA.
OC Q9X3P5;	
DT 01-NOV-1999 (TrEMBLrel. 12, Created)	
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE XYNA.	
GN XYNA.	
OS Caldicellulosiruptor sp. Tok7B.1.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC Thermoanaerobacter group; Caldicellulosiruptor.	
OX NCBI_TaxID=80339;	
RP [1]	
RC SEQUENCE FROM N.A.	
RA MEDLINE=20171169; PubMed=10706665;	
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,	
RA Bergquist P.L.;	
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme	
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";	
DR Curr. Microbiol. 40:333-340(2000).	
DR EMBL; AF078737; AAD30363.1; -	
DR HSSP; Q06851; INBC.	
DR InterPro; IPR001956; CBD_3.	
DR InterPro; IPR003305; CBD_6.	
DR InterPro; IPR001000; Glyco_hydro_10.	
DR InterPro; IPR002965; P_rich_extensn.	
DR Pfam; PF00942; CBD_3; 3.	
DR Pfam; PF02018; CBD_6; 2.	
DR Pfam; PF00331; Glyco_hydro_10; 1.	
DR PRINTS; PR00134; GLHYDRLASE10.	


```

Qy 64 MGCNTRASFSGSVNPATPTADTYLQ 88
   1 1 1 : : : : : : : : : :
Db 903 IGASNVTFNFVKLSSGVSGADYYLE 927

RESULT 12
O52374 PRELIMINARY; PRT; 1779 AA.
ID O52374;
AC O52374;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.8).
GN XYNL.
OS Caldicellulosiruptor sp. RT69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermocanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RT69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT RT69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF036924; AAB95326.1; -
DR HSP; Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 40.5%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 3.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps

Qy 4 GVKYQKNDSAPGDNQIKPGLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCDWAA 63
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1113 GLKVLKNNETSASTGSIRPWFKIVNGSSVDLSRVKIRIYWTVDGDKPQSAV-CDWAQ 1171

Qy 64 MGCNTRASFSGSVNPATPTADTYLQ 88
   1 1 1 : : : : : : : : : :
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196

RESULT 13
Q93TU6 PRELIMINARY; PRT; 499 AA.
ID Q93TU6;
AC Q93TU6;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE ALKALI TOLERABLE CELLULASE.
GN CEL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y106;
RA Qi M., Wang P., Liu X., Qu Y.;
RT "The cloning and expression of a cellulase from Bacillus subtilis
RT RT Y106.";
```


RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355629; AAK39540.1; -.
SQ SEQUENCE 499 AA; 55252 MW; 587A687584B7E5BE CRC64;

	Query Match	40.4%	Score 190;	DB 2;	Length 499;
	Best Local Similarity	40.7%;	Pred. No. 8.1e-13;		
	Matches	35;	Conservative	21;	Mismatches 26;
				Indels	4;
	Gaps				
Qy	4	GKVKYYKNNSGAPGDNIKPGQLVNTGGSSVDLSVTVRVWF--TRDGGSTLYVNCDA 62			
		: : : :	: : : :	: : : :	: : : :
		: : : :	: : : :	: : : :	: : : :
Db	353	GISVQYRAGDCGSMSNQIRPOLQIKNGNCTTDLKDVTARYWYAKAKKGQN--FDCDYA 409			
Qy	63	AMGCGRASFGSVNPATPTADTYLQ 88			
		: : : :	: : : :	: : : :	: : : :
Db	410	OIGCGNVTHKEVTLHKPKOGADTVLE 435			

RESULT	14
ID	Q93LDO
NAME	PRELIMINARY; PRT; 508 AA.
AC	Q93LDO;
DC	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxId=1423;
RN	[1]
RP	SEQUENCE FROM N. A.
STRAIN	CHZ1;
RA	Zvidrai C.J., Delgado O.D., Zvaunya R., Mattiasson B.;;
RC	"Bacillus subtilis CHZ1 endo-b-1,4-glucanase gene.";
RT	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AY044252; AAK94871.1; -
RR	Hydrolase; Glycosidase.
SK	SEQUENCE 508 AA; 56553 MW; A3955EE1C6F3340F2 CRC64;
QW	

Query Match	40.4%	Score 190;	DB 2;	Length 508;
Best Local Similarity	40.7%;	Pred. No. 8.3e-13;		
Matches	35;	Conservative	21;	Mismatches 26; Indels 4; Gaps 2;
Qy	4	GKVKYYNNNSGAPGDNQIKPGLQVLWNTGSSVDLSVTVRYWF	-TRDGGSTLVYNCDA	62
		:	:	:
		:	:	:
		:	:	:
		:	:	:
Db	362	GISVQYRAGDGSMSNQIRPOLQIKNGNTVLDKDVTRYWYKAKNGQN	--FDCDYA	418
Qy	63	AMGCNIRASFGSVNPATPTADTYLQ	88	
		:	:	:
		:	:	:
Db	419	OIGCGNVTHKFTVLHKPKOGADTYLE	444	

RESULT	ID	PRELIMINARY:	PRT:	499 AA.
OS	052731			
OC	052731;			
AC	052731;			
DT	01-JUN-1998 (TReMBLrel. 06, Created)			
DT	01-JUN-1998 (TReMBLrel. 06, Last sequence update)			
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)			
DE	ENDO-B-1,4-GLOUCANASE (EC 3.2.1.4).			
CELS.				
OS	Bacillus sp. 79-23.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=72363;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	SPRAIN-79-23;			
RC	Jung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;			
RC	"Cloning and expression of a Bacillus sp. 79-23 cellulase gene.";			
NT	Biotechnol. Lett 18:1077-1082(1996).			

RN	[2]
RP	SEQ
RC	STR
RA	YOO
RL	Sub
DR	EMB
DR	HSS
DR	Int
DR	Int
DR	Pfa
DR	Pfa
DR	Pro
DR	PRO
KW	Hyd
SQ	SEQ

	Query Match	39.8%	Score 187;	DB 2;	Length 499;
	Best Local Similarity	40.7%;	Pred. No. 1.8e-12;		
	Matches	35;	Conservative	20;	Mismatches 27; Indels 4; Gaps 2;
Qy	4	GKVVQYKNDSAPGDNQIKPGLQLVNTGSSVSVDLSVTVRYWF--TRDGGSGSTLYVNCDMA	62		
		: : :	:	:	:
		: : :	:	:	:
Db	353	GISVQYRAGDGSMSNQIRPOLQIKNGNNTVDLKDVTARYWYKAKNKQNV---DCDYA	409		
Qy	63	AMGCNTRASFSGSYNPATPTADTYLQ	88		
		: :	:	:	:
Db	410	QIGCGNTHKEVTLHKPKQGADTVLE	435		

Search completed: July 2, 2002, 09:17:34
Job time: 342 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:51 ; Search time 101.03 seconds
(without alignments)
96.748 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKVOYKKNDSAPGDNQ.....TRASFGSVNPATPTADTVLQ 88

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	231.5	49.4	782	AA15625	Cellulase AE-1. A
2	195.5	41.7	616	AAV13494	Truncated cellulase
3	194.5	41.5	1751	AAV13493	Truncated cellulase
4	191.5	40.8	1426	AAV13492	Truncated cellulase
5	186	39.7	499	AA12122	NK-1 cellulase. B
6	172	36.7	700	AA13227	Novel endoglucanase
7	165.5	35.3	551	AAW18790	Corrected Bacillus
8	165	35.2	167	AAW95080	Cellulose binding
9	165	35.2	476	AAV54123	A mannanase-linker
10	165	35.2	493	AAV28850	Pectate lyase-linker
11	165	35.2	493	AAV43218	Pectate lyase CBD

12	165	35.2	531	18	AAW15238	Scaffoldin protein
13	165	35.2	1853	19	AAW43108	C. thermocellum ce
14	155.5	33.2	1352	22	AAG63962	Amino acid sequenc
15	150.5	32.1	1350	22	AAG63963	Amino acid sequenc
16	120.5	25.7	531	16	AAW01503	60 kD endoglucanas
17	120.5	25.7	532	12	AA13229	Endoglucanase enco
18	114	24.3	162	15	AAE63634	Cellulose binding
19	114	24.3	162	20	AAW90077	C. cellulovorans C
20	114	24.3	163	22	AAE05745	Clostridium cellul
21	114	24.3	256	22	AAE81128	Cl17E2 OSPA constru
22	114	24.3	328	22	AAE05749	Chimeric S peptide
23	114	24.3	341	22	AAE05747	Clostridium cellul
24	114	24.3	428	22	AAE05748	Clostridium cellul
25	113	24.1	190	22	AAE05746	Clostridium cellul
26	113	24.1	382	20	AAV39952	Gaussia luciferase
27	112	23.9	154	20	AAW90081	C. cellulovorans C
28	112	23.9	156	20	AAW90080	C. cellulovorans C
29	75.5	16.1	986	21	AAV40440	Human brain-derive
30	75.5	16.1	1346	22	AAU04567	Human G-protein co
31	75.5	16.1	1346	22	AAU04581	Human G-protein co
32	75.5	16.1	1371	22	AAU04570	Novel human diagno
33	75	16.0	1723	22	ABG24680	Novel human respir
34	72.5	15.5	69	22	AAU17759	Fragment F1029 of
35	70.5	15.0	2965	19	AAW56450	Fragment HGJ1916 o
36	69	14.7	2970	19	AAW56445	Fragment HGJ1916 o
37	68	14.5	2873	19	AAW56441	Nascurium xyloglu
38	67.5	14.4	857	17	AAW5287	Sequence of the si
39	67	14.3	236	15	AAW45442	Anti-erbB2 scfv.
40	66	14.1	237	17	AAW4020	Single-chain anti-
41	66	14.1	237	18	AAW15185	Human p53 regulato
42	65.5	14.0	1566	20	AAV06309	Protein sequence o
43	64	13.6	117	20	AAW86130	Murine monoclonal
44	64	13.6	143	20	AAV50155	Drosophila melanog
45	64	13.6	254	22	ABB64503	

ALIGNMENTS

RESULT 1

AA15625

ID AA15625 standard; Protein; 782 AA.

AC AA15625;

XX 17-MAR-1992 (first entry)

DT Cellulase AE-1.

DE Detergents; pharmaceuticals; deinking; carboxymethylcellulose.

KW Aeromonas strain no. 212.

OS JP03251174-A.

PN 08-NOV-1991.

PD 28-FEB-1990; 90JP-0045465.

PF 28-FEB-1990; 90JP-0045465.

PR (OJIP) OJI PAPER KK.

PA WPI; 1991-373412/51.

XX N-PSDB; AAQ15178.

DR Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -

PT of opt. pH when carboxymethylcellulose is used as substrate.

PS Claim 2; Fig 3; 8pp; Japanese.

XX The sequence was deduced from the gene which was sequenced from

CC plasmid, pABC 1, prep. by ligating chromosomal DNA contg. the

CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
CC methylcellulose is the substrate. The N-terminal sequence: GIHADP-
CC has been confirmed by Edman degradation. The gene can be used to
CC produce recombinant enzyme which is used for the effective utilis-
CC action of biomass resources and the mfr. of pharmaceuticals and
CC foodstuffs, and also for the detergent and deinking of waste paper.
XX
XX
SQ Sequence 782 AA;

Query Match 49.4%; Score 231.5; DB 12; Length 782;
Best Local Similarity 48.3%; Pred. No. 5e-18;
Matches 42; Conservative 16; Mismatches 28; Indels 1; Gaps 1;

QY 2 SGGVKVQYKKNDSAPGDNQIKPGQLVNTGSSVDLSTVTVRVWFTRDGSSSLVYNCW 61
Db 631 sgdldavdyktgtdnaadnqfkphfnlrvnkgaavpilselslryyftad-gndqlqncdw 689
QY 62 AAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 690 amvgcslngafvkmnpghanadtyle 716

RESULT 2
AAY13494
ID AAY13494 standard; Protein; 616 AA.
XX
AC AAY13494;
XX
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulase Cel E3/B5.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
XX 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
XX (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI: 1999-315403/27.
DR N-PSDB; AAX55660.
XX
XX New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
XX Claim 7; Page 42-43; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or

CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.
XX
XX
SQ Sequence 616 AA;

Query Match 41.7%; Score 195.5; DB 20; Length 616;
Best Local Similarity 43.2%; Pred. No. 5.4e-14;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVKQYKKNDSAPGDNQIKPGQLVNTGSSVDLSTVTVRVWFTRDGSSSLVYNCW 60
Db 1 mgsgrkvlyknnetsastgsirpfwklvngsssvdlsrvkirywyvcdgdkpqsav-cd 59
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 60 waqigasvntfnfklssgvsgadyyle 87

RESULT 3
AAY13493
ID AAY13493 standard; Protein; 1751 AA.
XX
AC AAY13493;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulases comprising amino acid sequence.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
XX 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
XX (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI: 1999-315403/27.
DR N-PSDB; AAX55662.
XX
XX New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
XX Claim 7; Page 37-41; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic

Db 410 qicgcnthkfvtlhpkpgadyle 435

RESULT 6

AAAR13227
ID AAR13227 standard; Protein; 700 AA.

XX AC AAR13227;

XX DT 14-OCT-1991 (first entry)

XX DE Novel endoglucanase.

XX KW Cellulase activity; detergent.

XX OS Bacillus spp. NCIMB 40250.

XX FH Key Location/Qualifiers

FT Peptide 1..31

FT /label= signal sequence

FT Protein 32..700

FT /label= mature endoglucanase

FT Cleavage-site 31..32

XX PN WO9110732-A.

XX XX 25-JUL-1991.

XX XX 18-JAN-1991; 91WO-DK00013.

XX XX 19-JAN-1990; 90DK-0000164.

XX XX (NOVO) NOVO NORDISK A/S.

XX XX Jorgensen PL, Schulein M, Hansen C;

XX DR WPI; 1991-238020/32.

XX DR N-PSDB; AAQ13001.

XX PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
XX PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
XX PT fabrics.

XX PS Claim 1; Page 80; 96pp; English.

XX CC The enzyme is produced by a strain of Bacillus spp. NCIMB 40250
XX CC and exhibits an endoglucanase activity of at least 10 (pref. at
XX CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total
XX CC protein under alkaline conditions. It is especially useful as a
XX CC cellulolytic agent and has been found to be more stable during
XX CC washing (60 mins. at 40 deg.) in the presence of conventional
XX CC detergents than a commercial cellulase preparation. It may also
XX CC show increased storage stability in liq. detergents contg.
XX CC proteases. The sequence was deduced from the DNA (AAQ13001).
XX CC See also AAR13228 and AAR13229.

XX SQ Sequence 700 AA;

Query Match 36.7%; Score 172; DB 12; Length 700;

Best Local Similarity 44.9%; Pred. No. 3.3e-11;

Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKNNND-SAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWFTDRGSGSTLVYNC 59
| : : | | | | : | : | | | | : | : | | | | : | : | | | | : | : |

Db 549 vnsdlvvqykdgdrnatdqkphfnqkgtspvdlssitryftkd-ssaanngwi 607

QY 60 DWAMCGNIRASFGSVNPTPTADTYLQ 88

Db 608 dwaklgsgniqsfghnga--dsdtyae 634
| | : | : | | : | : | | : | : |

RESULT 7

AAW18790

ID AAW18790 standard; protein; 551 AA.

XX AC AAW18790;

XX DT 18-NOV-1997 (first entry)

XX DE Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.

XX KW Endoglucanase; Endo 3A; formation; localised; variation;
XX KW colour density; surface; dye; fabric; family 5; cellulose;
XX KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
XX KW blue jeans; back staining.

XX OS Bacillus lautus.

XX PN WO9709410-A1.

XX PD 13-MAR-1997.

XX PF 03-SEP-1996; 96WO-DK00364.

XX PR 08-SEP-1995; 95DK-0000993.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Fich M, Onishi M, Schulein M, Toft AH;

XX DR WPI; 1997-192888/17.

XX PT Localised variation of colour density in the surface of a dyed
XX PT cellulosic fabric - uses cellulase compsn. able to hydrolyse
XX PT p-nitrophenyl -beta-1,4-cellobioside

XX PS Disclosure; Pages 15-17; 23pp; English.

XX CC The present sequence is the corrected version of the incorrect
XX CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
XX CC WO9110732. Endo 3 can be used in novel method of forming localised
XX CC colour density variation on the surface of a dyed cellulosic
XX CC fabric. The method comprises agitating the fabric in an aqueous
XX CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
XX CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
XX CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
XX CC abrading agent or cellulose having abrading activity. Each
XX CC cellulase displays 30 % or more of its maximum activity at pH 7.
XX CC The process is useful to provide a stone washed look to blue jeans
XX CC without back staining.

XX SQ Sequence 551 AA;

Query Match 35.3%; Score 165.5; DB 18; Length 551;

Best Local Similarity 37.9%; Pred. No. 1.4e-10;

Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVOYKNNDSAPGDNQIKPGLQVLVNTGSSVDLSTVTYRYWFTDRGSGSTLVNCDW,61
| : | | | | : | : | | | | : | : | | | | : | : | | | | : | : |

Db 402 tgnlvvqykvgdtsatcdnqmkpsfnikngtptpnlsiglkryyftkd-gtdamsasfdw 460

QY 62 AAMCGNIRASFGSVNPTPTADTYLQ 88

Db 461 aqigasnvsaaf--aftsntdtyve 485
| : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8

AAAR95080

ID AAR95080 standard; peptide; 167 AA.

XX AC AAR95080;

XX DT 30-OCT-1996 (first entry)

XX XX

4 GVKVQYKNDSAPGDNQIKPGLQLVNTGSSSVDLSTVTVRYWFTRDGGSSTLVYNCDWAA 63

```
glvlqyrtadtnvndnlnphfqlnkgtisvpinelkirytyidgreq-tfncdyav
```

Qy 64 MCGNIRASFGSVNPATPTADTYLQ 88

```

Db 1258 lscsklngklvkmdkaatgadyytle 1282
      : | : :: | | | | |

```

Search completed: July 2, 2002, 09:15:52
Job time: 240 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:55 ; Search time 39.87 Seconds
(without alignments)
53.912 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKVOYKKNDSAPGDNQ.....IRASFGSYNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.7	616	4	US-09-136-574A-47
2	194.5	41.5	1751	4	US-09-136-574A-44
3	191.5	40.8	1426	4	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	493	4	US-09-198-956-10
8	165	35.2	493	4	US-09-198-955A-12
9	120.5	25.7	531	2	US-07-862-588B-7
10	114	24.3	162	1	US-08-048-164A-2
11	114	24.3	162	1	US-08-460-462-2
12	114	24.3	162	1	US-08-460-457-2
13	114	24.3	162	1	US-08-460-458-2
14	114	24.3	162	2	US-08-460-455-2
15	114	24.3	162	2	US-08-330-394A-2
16	114	24.3	163	3	US-09-006-636-7
17	114	24.3	163	4	US-09-006-632-7
18	113	24.1	382	4	US-09-277-716-22
19	112	23.9	154	2	US-08-330-394A-29
20	112	23.9	156	2	US-08-330-394A-22
21	64	13.6	428	3	US-09-118-319-5
22	64	13.6	484	1	US-08-353-400-36
23	63.5	13.5	1785	4	US-09-341-587-3
24	63	13.4	1581	4	US-09-110-517-2
25	62.5	13.3	288	4	US-09-423-439-38
26	62.5	13.3	445	1	US-08-353-400-33
27	62.5	13.3	673	4	US-09-423-439-32

28	62.5	13.3	802	4	US-09-081-345-18
29	62	13.2	1290	1	US-08-470-350B-2
30	61.5	13.1	128	1	US-07-946-421-26
31	61.5	13.1	307	2	US-08-484-905-68
32	61.5	13.1	307	3	US-08-481-985B-68
33	61.5	13.1	307	4	US-08-370-476-68
34	61.5	13.1	316	1	US-08-260-202A-18
35	61.5	13.1	316	1	US-08-017-114-18
36	61.5	13.1	316	3	US-08-505-307-18
37	61.5	13.1	316	5	PCT-US94-02034-18
38	61	13.0	301	2	US-08-656-906-25
39	61	13.0	301	4	US-09-217-847-25
40	61	13.0	617	1	US-08-361-920-29
41	61	13.0	617	1	US-08-479-939-29
42	61	13.0	617	1	US-08-483-432-29
43	60.5	12.9	240	1	US-08-488-113B-147
44	60.5	12.9	240	1	US-08-477-484B-147
45	60.5	12.9	240	2	US-08-646-360-147

ALIGNMENTS

RESULT 1
US-09-136-574A-47
; Sequence 47, Application US/09136574A-
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Sequence 18, Appl
Sequence 2, Appl
Sequence 26, Appl
Sequence 68, Appl
Sequence 68, Appl
Sequence 68, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 29, Appl
Sequence 29, Appl
Sequence 29, Appl
Sequence 147, App
Sequence 147, App
Sequence 147, App

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 4; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.2e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 60
Db 1 MGSGVKVLYKNNETSASTGSIKPFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 59
QY 61 WAAGCGNIRASFGSVNPTPTADTYLQ 88
Db 60 WAQIGASNVTFNEVKLSGSGVADYYLE 87

RESULT 2

US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44
Query Match 41.5%; Score 194.5; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 5.7e-13;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 63
Db 678 GVKVLYKNNETSASTGSIKPFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 736
QY 64 MGCNIRASFGSVNPTPTADTYLQ 88
Db 737 IGASNVTFNEVKLSGSGVADYYLE 761

RESULT 3

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43
Query Match 40.8%; Score 191.5; DB 4; Length 1426;
Best Local Similarity 43.5%; Pred. No. 9.5e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 63
Db 413 GLKVLKNNETSASTGSIKPFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 471
QY 64 MGCNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNEVKLSGSGVADYYLE 496

STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13813
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.2%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 5.7e-11;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTROGGSSTLVYVNC 60
Db 5 VSGNLKVEFVNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYRYVTVGDKQDTFW-CD 63
QY 61 WAAM-----GCCGIRASFGSVNPATPTADTYLQ 88
Db 64 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 100

RESULT 7
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.2%; Score 165; DB 4; Length: 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTROGGSSTLVYVNC 60
Db 331 VSGNLKVEFVNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYRYVTVGDKQDTFW-CD 389
QY 61 WAAM-----GCCGIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 8
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQLVNTGSSVDLSVTYVRYWFTROGGSSTLVYVNC 60
Db 331 VSGNLKVEFVNSPDTTNSINPQFKVTNTGSSAIDLSKLTLYRYVTVGDKQDTFW-CD 389
QY 61 WAAM-----GCCGIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 9
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7


```

RESULT 10
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

11
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELE: 66141 PENNIE

```


INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

Query Match 24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 2.3e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVVDLSTVTYRYWFTRDGSSSTLVYNCW 61

Db 3 TSSMSVEFYNSKSAQNTSITPIIKITNTSDSLNLDKVKRYVYTS DGTGGQTFW-CDH 61

QY 62 AMCGCN-----IRASF--GSVNPATPTADTYLQ 88

Db 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97

RESULT 12

US-08-460-457-2

; Sequence 2, Application US/08460457

; Patent No. 5719044

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; APPLICANT: Shpiegl, Itai

; APPLICANT: Goldstein, Marc A.

; APPLICANT: Dol, Roy H.

; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,457

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/048,164

; FILING DATE: 14-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7809-008

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 162 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-460-457-2

Query Match 24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 2.3e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVVDLSTVTYRYWFTRDGSSSTLVYNCW 61

Db 3 TSSMSVEFYNSKSAQNTSITPIIKITNTSDSLNLDKVKRYVYTS DGTGGQTFW-CDH 61

RESULT 13

US-08-460-458-2

; Sequence 2, Application US/08460458

; Patent No. 5738984

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAI

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/048,164

; FILING DATE: 14-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7809-007

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 162 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-460-458-2

Query Match 24.3%; Score 114; DB 1; Length 162;

Best Local Similarity 29.9%; Pred. No. 2.3e-05;

Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVVDLSTVTYRYWFTRDGSSSTLVYNCW 61

Db 3 TSSMSVEFYNSKSAQNTSITPIIKITNTSDSLNLDKVKRYVYTS DGTGGQTFW-CDH 61

RESULT 14

US-08-460-455-2

; Sequence 2, Application US/08460455

; Patent No. 5837814

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; APPLICANT: Shpiegl, Itai

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:14:14 ; Search time 54.67 Seconds
(without alignments)
154.671 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191	40.7	1711	2 T31337	1,4-beta-glucanase
2	190.5	40.6	1779	2 T31085	xylanase - Caldice
3	190	40.5	508	2 G69593	cellulase (EC 3.2.
4	190	40.5	508	2 A26874	cellulase (EC 3.2.
5	188	40.1	499	2 JN0111	cellulase (EC 3.2.
6	186.5	39.8	1331	2 A48954	mannan endo-1,4-be
7	186.5	39.8	1742	2 T17120	cellulase (EC 3.2.
8	186	39.7	499	2 A27198	cellulase (EC 3.2.
9	182.5	38.9	915	2 A43802	cellulase (EC 3.2.
10	182.5	38.9	1039	2 S02711	cellulase (EC 3.2.
11	180.5	38.5	145	2 A41897	cellulase homolog
12	180	38.4	486	2 I40548	bifunctional cellu
13	172	36.7	700	2 B41897	cellulase (EC 3.2.
14	165.5	35.3	879	2 A47704	endoglucanase I (E
15	165	35.2	1854	2 S36859	cipA protein - Clo
16	149.5	31.9	505	2 S39962	endoglucanase - Er
17	147.5	31.4	504	2 S54744	cellulase (EC 3.2.
18	141.5	30.2	986	2 S12021	thermoactive cellu
19	130.5	27.8	586	2 PC6006	scaffolding protei
20	126.5	27.8	1483	2 C97012	probably celluloso
21	126.5	27.0	1162	2 T30433	scaffolding protei
22	117	24.9	1230	2 S47466	cellulose 1,4-beta
23	114	24.3	1848	2 A44140	cellulose-binding
24	75.5	16.1	618	2 T08685	hypothetical prote
25	73.5	15.7	547	2 T23478	hypothetical prote
26	73.5	15.7	1428	2 AC2224	hypothetical prote
27	70	14.9	5188	2 B85547	probable RTX famil
28	70	14.9	5291	2 F90696	hypothetical prote
29	68.5	14.6	574	2 A69196	cell surface glyco

30	67	14.3	271	2 H72684	hypothetical prote
31	66.5	14.2	247	2 JC7201	thumatin-like pro
32	66	14.1	535	2 T47790	hypothetical prote
33	65.5	14.0	751	2 T40462	ser-lys rich hypot
34	65.5	14.0	1435	2 T01075	polyprotein - hepa
35	65	13.9	523	2 G70717	probable purH prot
36	64	13.6	1222	2 G72614	probable reverse g
37	64	13.6	1571	2 S50669	hypothetical prote
38	64	13.6	1594	2 T30549	hensin - rabbit
39	63.5	13.5	398	2 H70393	hemolysin - Aquife
40	63.5	13.5	2403	2 A59386	sanko - human
41	62.5	13.3	269	2 H95976	endo-beta-1,3-1,4-
42	62.5	13.3	454	2 AB2512	replicative DNA he
43	62.5	13.3	802	1 B44390	protein-tyrosine-p
44	62.5	13.3	4936	2 AH2515	hypothetical prote
45	62	13.2	122	2 E37267	Ig heavy chain v r

ALIGNMENTS

RESULT 1

T31337

1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C:Species: Anaerocellum thermophilum

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31337

R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144, 457-465, 1998

A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) f

omains.

A:Reference number: Z21003; MUID:98154434

A:Accession: T31337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1711 <ZVF>

A:Cross-references: EMBL:286105; NID:e1071329; PID:e350354; PIDN:CAB06786.1

C:Genetics:

A:Gene: celA

C:Keywords: glycosidase; hydrolase

Query Match 40.7%; Score 191; DB 2; Length 1711;

Best Local Similarity 43.8%; Pred. No. 1.2e-11;

Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-VKVOYKNNDSAPGDNQIRPGLQLVNTGSSVDLSTVTYVYVTRDGGSTLVYNC 59

Db 683 VAGGQIKVLYANKETSTNTIRPWLKXVNTGSSDLSLSTVIRYWTVDGDKAQSAIS- 741

QY 60 DWAAACGNNIRASFGSVNPATPTADTYLQ 88

Db 742 DWAGIGASNVTFKVLSSSVSGADYYLE 770

RESULT 2

T31085

xylanase - Caldicecellulosiruptor sp.

C:Species: Caldicecellulosiruptor sp.

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL Data Library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldicecellulosiruptor sp. Rt69B.1.

A:Reference number: Z20972

A:Accession: T31085

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 <MOR>

A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1

C:Genetics:

A:Note: xync

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 1.4e-11;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSSTLVYNCDDAA 63
DB 1113 GLKVLVKNNETSASTGSRPFKIVNGSSVDLSRKYRYWTYDGDGKPKQSAV-CDWAQ 1171
QY 64 MCGGNIRASFSGVNPATPTADTYLQ 88
DB 1172 IGASNVTFNFVKLTSGVSGADTYLE 1196

RESULT 3
cellulase (EC 3.2.1.4) bglc precursor - Bacillus subtilis
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C:Accession: G69593; S24239; S49103; I39803
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, A.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chikara, A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: G69593
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KUN>
A:Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CABI3696.1; PID:el183471;
R:Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Sell
Nucleic Acids Res. 14, 9159-9170, 1986
A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:87066783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 1-508 <MAC>
A:Cross-references: GB:X04689; NID:g39823; PIDN:CAA28392.1; PID:g39824
A:Experimental source: strain PAP115
A:Note: part of this sequence, including the amino end of the mature form, was confirmed
R:Lindahl, V.; As, K.; Trousmo, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
A:Reference number: I40353; MUID:95225656
A:Accession: I40353
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-508 <LIN>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
R:Lindahl, V.; As, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 1-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
A:Experimental source: strain CK-2
R:Wolf, M.; Geczi, A.; Borriass, R.
submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons

A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 1-508 <WOI>
A:Cross-references: EMBL:229076; NID:g509266; PIDN:CAA82317.1; PID:g509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
Korean J. Microbiol. 24, 236-242, 1986
A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
A:Reference number: I39803
A:Accession: I39803
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-508 <WOI>
A:Cross-references: GB:M38634; NID:g142657; PIDN:AAA22300.1; PID:g142658
A:Experimental source: strain ATCC 6633
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Genetics:
A:Gene: bglc
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Pathway: cellulose degradation
A:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.5%; Score 190; DB 2; Length 508;
Best Local Similarity 40.7%; Pred. No. 4.3e-12;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSSTLVYNCDDAA 62
DB 362 GISVQYRAGDGSMSNIRPQLQIKNGNTVDLKDVTARYWKAKNGQN---FDCDYA 418
QY 63 AMCGNIRASFSGVNPATPTADTYLQ 88
DB 419 QIGCGNVTHKFVTLHKPKQGADTYLE 444

RESULT 4
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB>
A:Cross-references: GB:M16185; NID:gl43007; PIDN:AAA22496.1; PID:gl43008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A:Note: the authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Pathway: cellulose degradation
A:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; po
F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.5%; Score 190; DB 2; Length 508;
Best Local Similarity 43.0%; Pred. No. 4.3e-12;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSSTLVYNCDDAA 62

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.8%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSSDVLSTVTVRWFTRDGGSSTLVYNCDW 61
|| || || : : | : || : || || || || || || || : ||
Db 364 SGQIKVLVYANKETNSTTIRPWLVKVVNSGSSDLSRVTIRYWTVTDGERAQSAIS-DW 422

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
| : | | : : | : || || :
Db 423 AQIGASNVTFFVKLLSSVSGADYYLE 449

RESULT 7
Tl7120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C;Species: Caldocellum saccharolyticum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: Tl7120; A43745
R;Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A;Title: CelA, another gene coding for a multidomain cellulase from the extreme thermophilic bacterium *Caldococcus*
A;Reference number: Z18698; MUID:95336703
A;Accession: Tl7120
A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A;Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R;Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A;Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding for a cellulase from the extreme thermophilic bacterium *Caldococcus*
A;Reference number: A43745; MUID:91247819
A;Accession: A43745
A;Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A;Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C;Genetics:
A;Gene: celA
C;Keywords: glycosidase; hydrolase

Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSSDVLSTVTVRWFTRDGGSSTLVYNCDW 61
|| || || : : | : || : || || || || || || || : ||
Db 704 SGQIKVLVYANKETNSTTIRPWLVKVVNSGSSDLSRVTIRYWTVTDGERAQSAIS-DW 762

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
| : | | : : | : || || :
Db 763 AQIGASNVTFFVKLLSSVSGADYYLE 789

RESULT 8
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Bacillus subtilis
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C;Accession: A27198
R;Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A;Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.
A;Reference number: A27198; MUID:87190397
A;Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAK>
A;Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671

A:Experimental source: strain IFO3034

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose; cellulose degradation
A:Pathway: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 39.7% Score 186; DB 2; Length 499;
Best Local Similarity 40.7% Pred. No. 1.1e-11;
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

Qy 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 353 GISVQYKAGDGVNSQIRPOLHIKNKGNATVLDKDTARYWYNAKNQGN---FDGDAY 409

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 410 QIGCGNLTHKEVTLHKPKQGADTYLE 435

RESULT 9

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum s

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Caldocellum saccharolyticum

C>Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998

C:Accession: A43802

R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "

A:Reference number: A43802; MUID:91136262

A:Accession: A43802

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A:Molecule type: DNA

A:Residues: 1-915 <SAU>

A:Cross-references: EMBL:X13602

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.9% Score 182.5; DB 2; Length 915;

Best Local Similarity 41.9% Pred. No. 4.8e-11;

Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 349 GQIKVLYANKETNSTTIRPWLKVVNSGSSIDLSTLRITRYWTVVDGERAQSAYS-DWA 407

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 408 QIGASNTTFKFKLVSSVSGADTYLE 433

RESULT 10

S02711

cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Caldocellum saccharolyticum

C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999

C:Accession: S02711

R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.

Nucleic Acids Res. 17, 439, 1989

A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for ex

A:Reference number: S02711; MUID:89098398

A:Accession: S02711

A:Molecule type: DNA

A:Residues: 1-1039 <SAU>

A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646

C:Genetics:

A:Gene: celB

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-28/Domain: signal sequence #status predicted <SIG>

F:23-1039/Product: cellulase #status predicted <MAT>

F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.9% Score 182.5; DB 2; Length 1039;

Best Local Similarity 41.9% Pred. No. 5.6e-11;

Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 420 GQIKVLYANKETNSTTIRPWLKVVNSGSSIDLSTLRITRYWTVVDGERAQSAYS-DWA 478

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 479 QIGASNTTFKFKLVSSVSGADTYLE 504

RESULT 11

A41897

cellulase homolog - Bacillus lautus (fragment)

C:Species: Bacillus lautus

C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999

C:Accession: A41897; S27498

R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A:Title: celA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-

A:Reference number: A41897; MUID:92276330

A:Accession: A41897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <HAN>

A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662

A:Experimental source: PL236

A>Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 38.5% Score 180.5; DB 2; Length 145;

Best Local Similarity 43.9% Pred. No. 1.1e-11;

Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

Qy 7 VQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 66

Db 1 LQYRAADTNAADNQIKPSFNKNGTSAVDLSTLRITRYWTVVDGERAQSAYS-IDWAQIGG 59

Qy 67 GNTRASFSGSVNPATPTADTYLQ 88

Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 12

I40548

bifunctional cellulase precursor - Bacillus sp.

C:Species: Bacillus sp.

C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999

C:Accession: I40548

R:Han, S.J.; Yoo, Y.J.; Kang, H.S.

J. Biol. Chem. 270, 26012-26019, 1995

A:Title: Characterization of a bifunctional cellulase and its structural gene: the ce

A:Reference number: I40548; MUID:96029707

A:Accession: I40548

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-486 <RES>

A:Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576

Qy 61 WAAM-----CGGNIRASFGSVNPATPTADTYLQ 88
||:
Db 424 HAAIGSNGSYNGITSNVKGTFFVKMSSTNNADTYLE 460

Search completed: July 2, 2002, 09:14:14
Job time: 142 sec

[illegible]

RT bacterium 'Caldocellum saccharolyticum';
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L32742; AAA91086.1; -
CC EMBL: M35063; AAA72860.1; -
CC EMBL: L01257; -; NOT_ANNOTATED_CDS.
CC PIR: A43745; A43745.
CC HSSP: P26221; ITF4.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR000556; Glyco_hydro_48.
CC InterPro: IPR001701; Glyco_hydro_9.
CC Pfam: PF00942; CBD_3; 3.
CC Pfam: PF02011; Glyco_hydro_48; 1.
CC Pfam: PF00759; Glyco_hydro_9; 1.
CC PRINTS: P00844; GLHYDRLASE48.
CC ProDom: PD001947; CBD_3; 2.
CC ProDom: PD011903; Glyco_hydro_48; 1.
CC ProSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC ProSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
KW SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 39.8%; Score 186.5; DB 1; Length 1742;
Best Local Similarity 42.5%; Pred. No. 4.9e-13;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDW 61
DB 704 SGIKVLVYANKETNSTTIRPWLKVVNSGSSIDLRSVTRYWYTVYDGERAQAIS-DW 762
QY 62 AAMCGGNIRASFGSVNPATPTADYILQ 88
DB 763 AQIGASVNTFKVKLSSSVSGADYILE 789
RESULT 6
GUNE_CALSA STANDARD; PRT; 1039 AA.
ID GUNE_CALSA
AC P10474;

DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellobiohydrolase and endocellobiohydrolase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13602; CAA31936.1; -
CC PIR: S02711; S02711.
CC HSSP: Q06851; INBC.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001000; Glyco_hydro_10.
CC InterPro: IPR001547; Glyco_hydro_F5.
CC Pfam: PF00942; CBD_3; 1.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00331; Glyco_hydro_10; 1.
CC PRINTS: P00134; GLHYDRLASE10.
CC ProDom: PD001947; CBD_3; 1.
CC ProSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
CC ProSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat;
KW Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.
FT DOMAIN 376 416 THR/PRO-RICH, TANDEM REPEATS OF T-P.
FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 571 618 THR/PRO-RICH, TANDEM REPEATS OF T-P.
FT ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
FT ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 792 792 BY SIMILARITY.
SQ SEQUENCE 1039 AA; 117641 MW; 0B0378171594DDAE CRC64;

Query Match 38.9%; Score 182.5; DB 1; Length 1039;
Best Local Similarity 41.9%; Pred. No. 7.5e-13;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDA 62
DB 420 GQIKVLVYANKETNSTTIRPWLKVVNSGSSIDLRSVTRYWYTVYDGERAQAIS-DWA 478


```
CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39788; AAC37033.1; -.
DR HSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE DEGRADATION; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 444;
Best Local Similarity 38.6%; Pred. No. 1.6e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQKNDSPAGNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLYNC 59
DB 357 TGDVVLQYRNVDDNPSDDAIRMAVNIKTGSTPIKLSDLQVRYFYHDDGKPGANLFV--- 413
QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87
DB 414 DWANVGPNIVTGTGTPAASTDKANRYV 441

RESULT 13
GUNW_ERWCA STANDARD; PRT; 505 AA.
AC Q47096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
GN CELV.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmon G.P.C.;
RT "Molecular analysis of the major cellulase (Celv) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76000; CAA53592.1; -.
DR
```

```
DR HSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE DEGRADATION; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 505 ENDOGLUCANASE V.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 505;
Best Local Similarity 38.6%; Pred. No. 1.8e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQKNDSPAGNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLYNC 59
DB 354 TGDVVLQYRNVDDNPSDDAIRMAVNIKTGSTPIKLSDLQVRYFYHDDGKPGANLFV--- 410
QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87
DB 411 DWANVGPNIVTGTGTPAASTDKANRYV 438

RESULT 14
GUNW_ERWCA STANDARD; PRT; 504 AA.
AC Q59355;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
DE (Cellulase V1).
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79241; CAA55823.1; -.
DR HSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.
```


Search completed: July 2, 2002, 09:18:05
Job time: 373 sec

Result No.	Query 8			DB	ID	Description
	Score	Match	Length			
1	225.5	48.1	170	2	Q9RFX6	Q9rfx6 caldibacill
2	216.5	46.2	930	2	Q9RFX5	Q9rfx5 caldibacill
3	214.5	45.7	921	2	Q9RL8L	Q9rl8l caldibacill
4	204.5	43.6	997	2	Q92411	Q92411 bacillus spp
5	194.5	41.5	1751	2	Q9AQG4	Q9aqg4 caldicellu
6	192.5	41.0	1000	2	Q24820	Q24820 thermophili
7	192.5	41.0	1770	2	Q9X3P5	Q9x3p5 caldicellu
8	191.5	40.8	261	2	Q9AQG7	Q9aqg7 caldicellu
9	191.5	40.8	1426	2	Q9X3P6	Q9x3p6 caldicellu
10	191	40.7	1711	2	P96311	P96311 anaerocellu
11	190.5	40.6	996	2	Q9AQH0	Q9aqh0 caldicellu
12	190.5	40.6	1779	2	Q52374	Q52374 caldicellu
13	190	40.5	499	2	Q93TJ6	Q93tj6 bacillus su
14	190	40.5	508	2	Q93LD0	Q93ld0 bacillus su
15	187	39.9	499	2	Q52731	Q52731 bacillus spp
16	186	39.7	499	2	Q45532	Q45532 bacillus su

DE	BETA-GLUCANASE.
OS	thermophilic anaerobe NA10.
OC	Bacteria.
OX	NCBI_TaxID=67756;
EN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=NA10;
RA	Miyake K., Machida Y., Hattori K., Iijima S.;
RT	"Characterization of a multi-domain cellulase from an extremely thermophilic anaerobe strain NA10.";
RL	Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AB008029; BAA22939.1; -
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001589; Actinin_act_bind.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR001000; Glyco_hydro_10.
DR	InterPro; IPR001547; Glyco_hydro_F5.
DR	Pfam; PF00942; CBD_3; 1.
DR	Pfam; PF00150; cellulase; 1.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS; PR00134; GLHYDLASE10.
DR	PRODOM; PD001947; CBD_3; 1.
DR	PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR	PROSITE; PS00591; GLYCOSYL_HYDROL_E10; 1.
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_N_1.
SQ	SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match	41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity	43.7%; Pred.No. 9.9e-13;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps	

Qy	2	SGGVKVOYKNNDAPGDNQIKPGIQLVNTGSSSVDLSTVTTRYWFTRDGGSGTLVYNCDW 61
		: : : : : : : : : : : : : :
Db	372	SGQIKVLANKETNSTTIRPWLVKVNTGSSSIDLSRVTIRVYWTVDGDKQAQSAVS-DW 430

Qy	62	AAMCGGNIRASFGSVNPATPTADTYLQ 88
		: : : : : : : : : : : :
Db	431	AQIGASNVTFRFKVLSSVSGADYYLE 457

RESULT	7
ID	Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC	Q9X3P5;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	XYNA.
GN	XYNA.
OS	Caldicellulosiruptor sp. Tok7B.1.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Thermoanaerobacter group; Caldicellulosiruptor.
OX	NCBI_TaxID=80339;
EN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=TOK7B.1;
RX	MEDLINE=20171169; PubMed=10706665;
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.;
RT	Bergquist P.L.;
RT	"Multi-domain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL	Curr. Microbiol. 40: 333-340(2000).
DR	EMBL; AF078737; AAD30363.1; -
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR003305; CBD_6.
DR	InterPro; IPR001000; Glyco_hydro_10.
DR	InterPro; IPR002965; Prich_extensn.
DR	Pfam; PF00942; CBD_3; 3.
DR	Pfam; PF02018; CBD_6; 2.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS; PR00134; GLHYDLASE10.


```

QY      64 MGCNINIRASFGSVNPATPTADTYLQ 88
       :| | :| | :| |
Db     903 IGASNVTFNFVKLSGGVSGADYYLE 927

RESULT 12
O52374 PRELIMINARY; PRT; 1779 AA.
AC
DT 01-JUN-1998 (T+EMBLrel_06, Created)
DT 01-JUN-1998 (T+EMBLrel_06, Last sequence update)
DT 01-DEC-2001 (T+EMBLrel_19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.18).
GN XYN.
OS Caldicellulosiruptor sp. Rt69B.l.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RC STRAIN=RT69B.l.;
RP MORRIS D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.l.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RR EMBL; AF036924; AAB95326.1; -.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBD_3; 3.
DR Pfam; PF02018; CBD_6; 2.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDLASE10.
DR PRODOM; PD001947; CBD_3; 3.
DR PROSITE; PS00591; GLYCOSYL-HYDROL_F10; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 3.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY      4 GVKQYKNNDSAPGDNOIKPGQLQIVNTGSSVDLSLTVTYRYWFTRDGGSSTLVNCWAA 63
       :||| ||||::: |:| :|| ||||||| :|||: || ||||
Db    1113 GLKVLYKNNETSASTGSRPFWKIVNGSSVDLSLRKIRYIYTVGDKPOSAY-CDWAQ 1171

QY      64 MGCNINIRASFGSVNPATPTADTYLQ 88
       :| | :| | :| |
Db    1172 IGASNVTFNFVKLSGGVSGADYYLE 1196

RESULT 13
Q93TJ6 PRELIMINARY; PRT; 499 AA.
AC
DT 01-DEC-2001 (T+EMBLrel_19, Created)
DT 01-DEC-2001 (T+EMBLrel_19, Last sequence update)
DT 01-DEC-2001 (T+EMBLrel_19, Last annotation update)
DE ALKALI TOLERABLE CELLULASE.
GN CEL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Y106;
RA Qi M., Wang P., Liu X., Qu Y.;
RT "The cloning and expression of a cellulase from Bacillus subtilis
   RT Y106.";
```


THIS PAGE BLANK (USPTO)